



This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book's long journey from the publisher to a library and finally to you.

### Usage guidelines

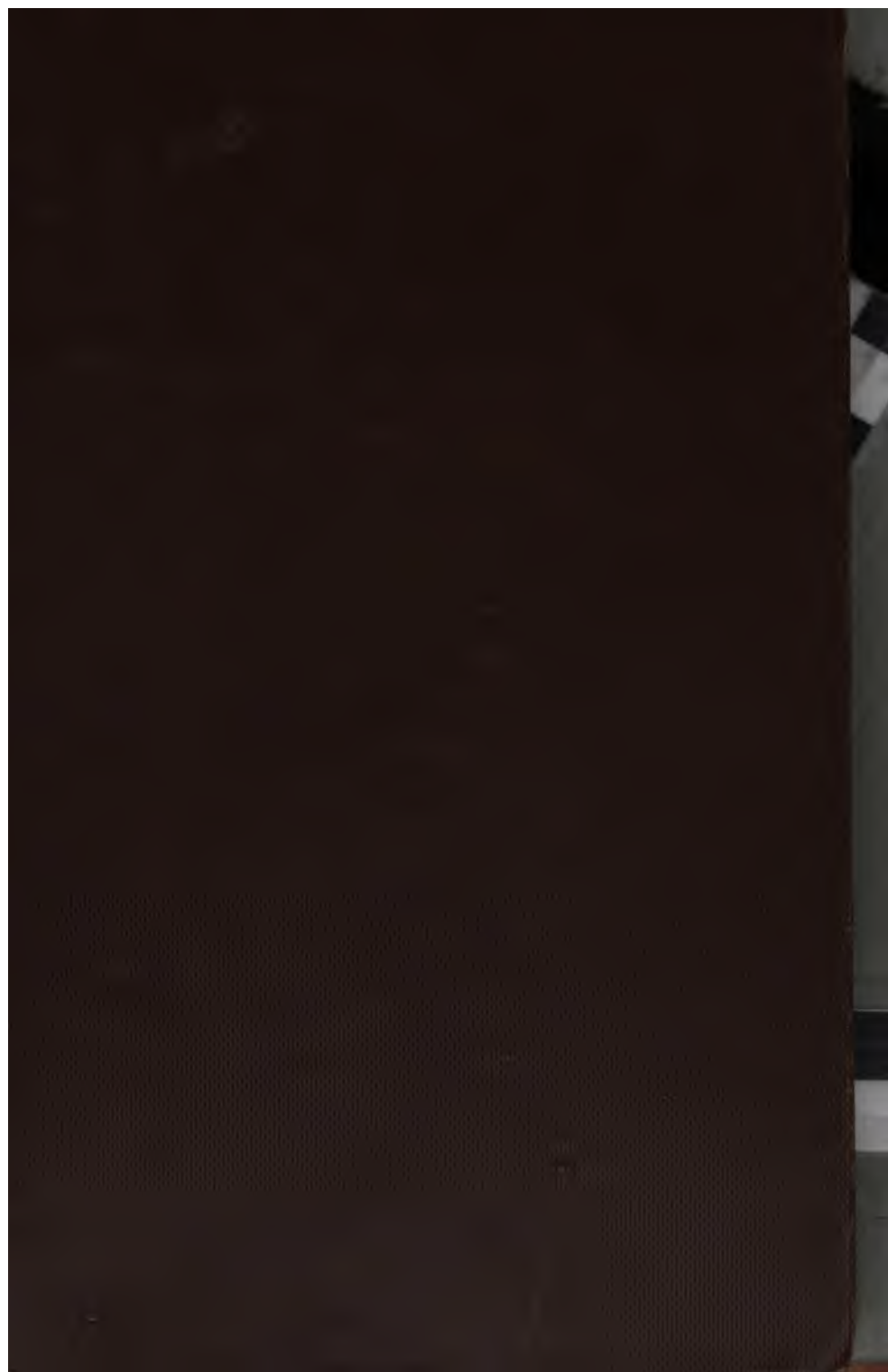
Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

- + *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.
- + *Refrain from automated querying* Do not send automated queries of any sort to Google's system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.
- + *Maintain attribution* The Google "watermark" you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.
- + *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can't offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book's appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

### About Google Book Search

Google's mission is to organize the world's information and to make it universally accessible and useful. Google Book Search helps readers discover the world's books while helping authors and publishers reach new audiences. You can search through the full text of this book on the web at <http://books.google.com/>



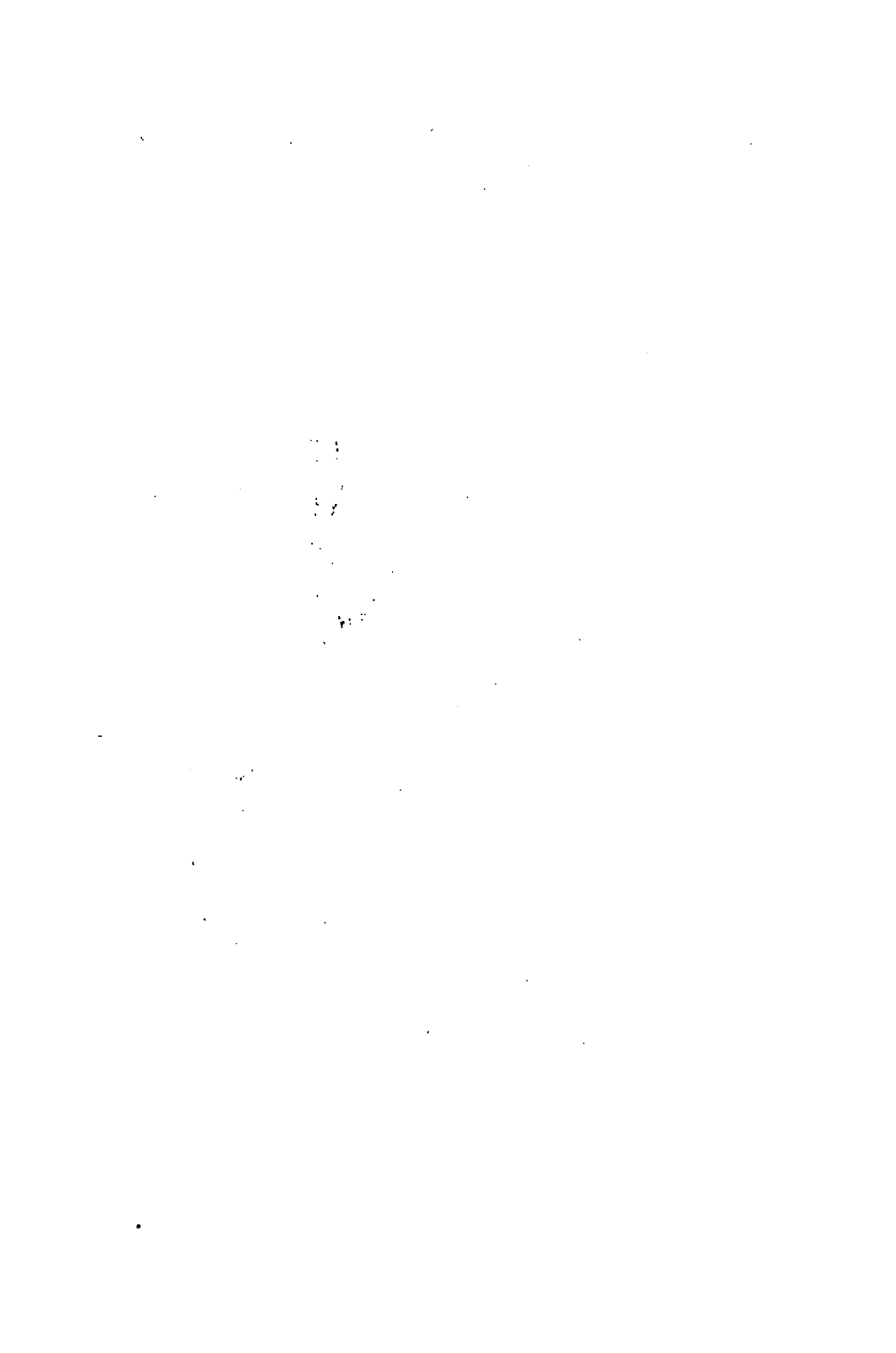
371.27  
W748  
cop 1



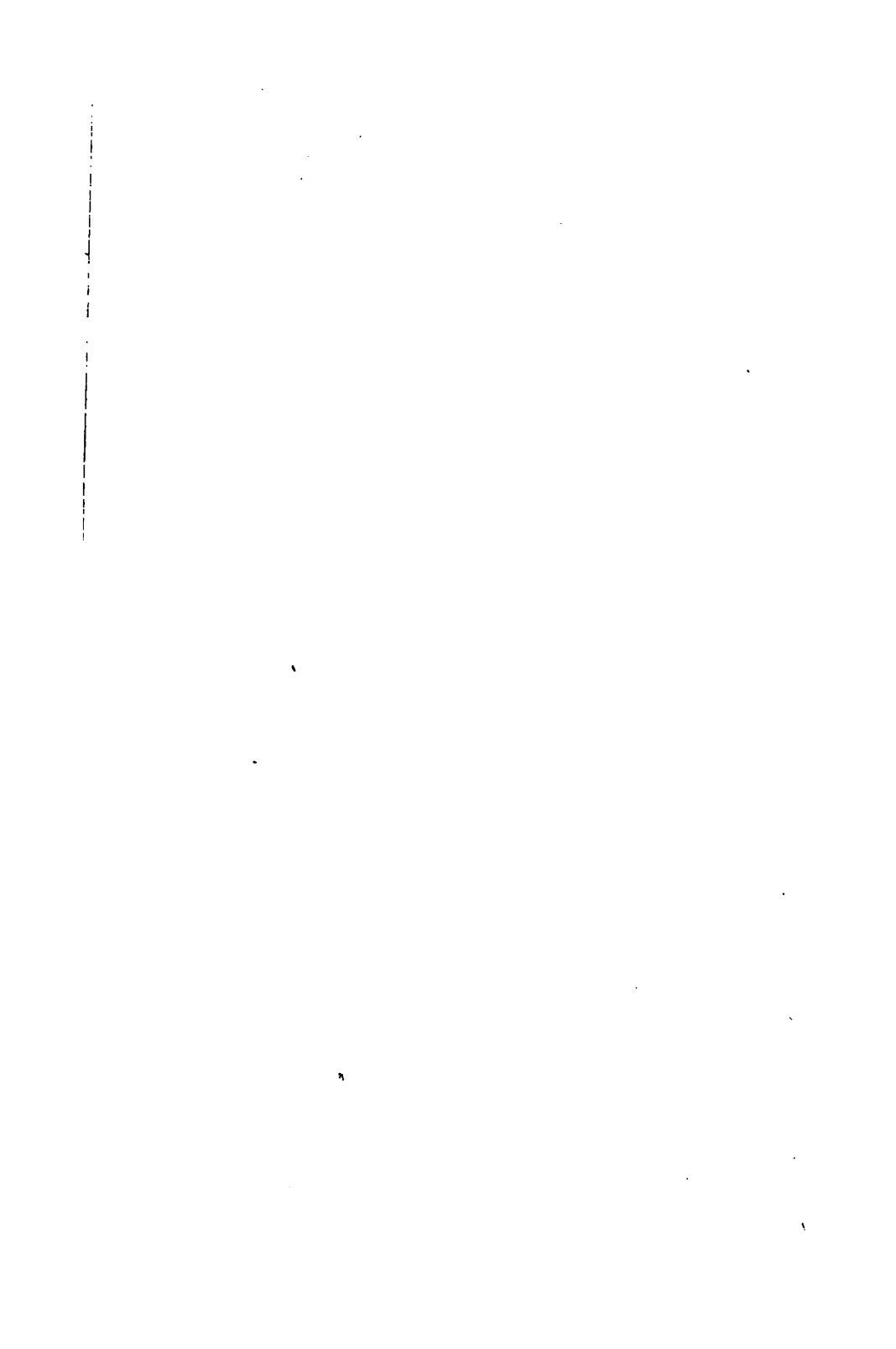
LELAND STANFORD JUNIOR UNIVERSITY



SCHOOL OF  
LIBRARY SCIENCE







## HOW TO MEASURE



THE MACMILLAN COMPANY  
NEW YORK • BOSTON • CHICAGO • DALLAS  
ATLANTA • SAN FRANCISCO

MACMILLAN & CO., LIMITED  
LONDON • BOMBAY • CALCUTTA  
MELBOURNE

THE MACMILLAN CO. OF CANADA, LTD.  
TORONTO

# HOW TO MEASURE

BY

G. M. WILSON, Ph.D.

PROFESSOR OF VOCATIONAL EDUCATION AND  
DIRECTOR OF THE SUMMER SESSION  
IOWA STATE COLLEGE OF  
AGRICULTURE AND MECHANIC ARTS  
AMES, IOWA

AND

KREMER J. HOKE, Ph.D.

DEAN AND PROFESSOR OF EDUCATION  
COLLEGE OF WILLIAM AND MARY  
WILLIAMSBURG, VIRGINIA  
FORMERLY SUPERINTENDENT OF PUBLIC SCHOOLS  
DULUTH, MINNESOTA

NEW YORK

New York

THE MACMILLAN COMPANY

1920

*All rights reserved*



## **HOW TO MEASURE**

Even so, when it was proposed at the Philadelphia meeting of the department of superintendence in 1913 that a committee on school efficiency be appointed, there was vigorous opposition. The proposal was merely for the appointment of a committee, yet a decision required a standing vote and carried by a majority of only one. The next year, at the Richmond meeting of the department of superintendence, it was surprising to note the change in sentiment.

The growth that may take place with an individual in a single year is well illustrated by the remarks of Superintendent Ben Bluett, of the St. Louis public schools. At the Philadelphia meeting, in his usual sincere and thorough way of thinking, he was very much disturbed that a group of young men should propose the measurement of "childhood," "mother love," and other intangible elements of the educative process. There was, in fact, never any intention of trying to measure these elements, but such terms were used by the opposition, and it was Ben Bluett's impassioned appeal against such procedure which had much to do with the large vote against the proposal for the appointment of a committee on measurement and school efficiency. A year later it was generally agreed that the feature of the Richmond meeting was Ben Bluett's confession. He had been made a member of the committee appointed at Philadelphia. He had met with this committee, fifteen in number, several times during the year, and had studied the question earnestly with the other members of the committee. He had begun to realize the significance of the movement and had secured the coöperation of Dr. Withers of the St. Louis College for Teachers in applying some of the tests in the St. Louis schools. The loyal, sincere, whole-hearted manner in which Ben Bluett acknowledged his lack of understanding of the movement a year before, and his thorough conversion to the advantages of the movement, swept away whatever opposition there may have been in the Richmond

meeting. From that time forward the progress of the movement has been only a question of ways and means, and better adaptation to secure the desired results. Even the school survey movement, that phase of the school efficiency movement which has been most feared by superintendents because of its frequent use by an opposition to discredit the work of the schools, has entered upon new life and has become an integral part of the American public school system.

It must not be assumed, however, that the work in measurement in the public schools has been perfected. It has passed the first stages. Leaders are convinced. Useful scales and tests have been developed. The technique of formulating a test has been further perfected and the value of a scientific test is better understood. In some respects we have entered the second stage of measurement. We have come to the point of discriminating between good and bad tests. Already a few standardized tests have been discarded.

We are now quite surely approaching a third stage of development, and that is the stage in which the tests shall be thoroughly weighed and judged as to the fundamental considerations of curricula making involved, whether they are or are not testing desirable school products, and whether their use will or will not lead to better methods of teaching and better selection of subject matter. In this stage the standard tests will be used more and more for the diagnosis of the weaknesses of individual pupils, more and more in testing the efficiency of methods of teaching. It is in this third stage that the rank and file of the teaching profession are necessarily involved. If the tests are to be of service, not merely as a general measure of the efficiency of a school system, but also of service *to the teacher* and *for the pupils* in the schoolroom, then it becomes necessary that the individual teacher shall master the details for actually using the tests in her own schoolroom. This is not too much to expect if a



man well beyond sixty, as was Superintendent Bluett, could approach this movement with an open mind and accept its benefits after a year of conscientious study.

That teachers are interested and keen to master the accumulated knowledge with regard to measurement is more and more apparent. Hence this effort is made to bring together the various contributions on the subject in form for use by the teacher. It is true, of course, that we shall make slow progress in educating the entire teaching profession until teachers become a trained body of professional educators with permanent tenure. But for this it were unwise to wait. In the meantime, may we not expect that any one who has accepted the responsibilities of the teaching profession will consider that she owes it to herself and to her pupils to master the details of using scales and standardized tests for the measurement of subject matter?

## CHAPTER II

### THE MEASUREMENT OF SPELLING

THERE are at present several spelling tests available. Before deciding on which one to select for use, it will be well to consider what should be tested in spelling.<sup>1</sup>

It appears that a person needs to spell only when he writes. People are therefore good spellers, for all social purposes, when they spell correctly the words which they use in their written work, such as writing letters, articles, club papers, compositions, school exercises, business notes, and the like. Manifestly the words used under such circumstances are the foundation words of the English language. The first requirement of a test in spelling, therefore, is that it be based upon the common fundamental words of the English language.

**What to Test.** — Much progress has been made in determining the fundamental words in the English language. Dr. W. Franklin Jones, at the University of South Dakota, studied the writing vocabulary of grade pupils by analyzing the words in the composition work of 1050 pupils residing in four different states. The work was so managed as to lead pupils to cover all the various fields of experience, and so exhaust the words in their several vocabularies. The pupils continued to write until new words ceased to appear in their compositions. In all, 75,000 themes were secured, consisting

<sup>1</sup> The teacher who wants further help on the value of measurement in education should take time to read Chap. XII before proceeding with the present chapter. The teacher unfamiliar with statistical terms will need to consult Chap. XI as terms occur in this and succeeding chapters. For the practical uses of the spelling tests, see the last section of this chapter, beginning on page 19.

of a total of 15,000,000 words. Dr. Jones spent eight years collecting and scoring these data. When completed, it was found that a total of only 4532 different words had been used by all these pupils. The largest single vocabulary consisted of 2812 words, the vocabulary of an eighth grade girl. The result of this study was to give a list of words which accurately represents the fundamental words used by school children. Apparently, it contains also the fundamental words of the English language.

Other studies have been made. One of similar character, which has led to the formation of a spelling scale, was conducted by Dr. Leonard P. Ayres. Dr. Ayres examined a total of 368,000 words written by 2500 different persons. This was a summary of previous studies. The first of these studies included in all about 100,000 words taken from standard literary selections. The second was an analysis of 250 different articles which appeared in four Sunday newspapers published in Buffalo. The third consisted of the tabulations of 23,629 words from 2000 short business letters. The fourth consisted of some 200,000 words taken from the family correspondence of 13 adults.

The Ayres study has the advantage of being based upon the words used by adults, and if we assume that the schools must prepare for active social participation on the adult level, then certainly Dr. Ayres' study would be above criticism from the standpoint of determining the fundamental words of the English language in common use for writing purposes. The Jones study and the Ayres study are in complete agreement as to the simplicity and small compass of the writing vocabulary.

Any adequate test must be based upon the words of the language that are in common use and fundamental in written work.

**The Ayres Scale.** — In undertaking to form a scale for testing the spelling of school pupils, the first thing which

100  
100

of a t  
collec  
found  
used  
consis  
girl.  
which  
school  
menta

Oth  
which  
condu  
a tot  
This  
studie  
stand  
250 d  
paper  
tabul  
The f  
famil

Th  
the w  
must  
level,  
from  
of th  
The  
ment  
vocal

An  
lang  
work  
Th  
i

Dr. Ayres did was to determine the words which were *most* fundamental. The 368,000 words of his study were made up largely of repetitions. Fifty different words were repeated so frequently that they made up approximately half of the entire list. Dr. Ayres had fixed upon 1000 words as the number which he should select. In order to get the 1000 words, he finally took all words which had been repeated as many as 44 times in the entire study.

The next step was to arrange the different words according to difficulty, in order to secure a graded test, or, in other words, a *spelling scale*. To determine the relative difficulty of the words in the 1000 list, Dr. Ayres arranged to have the words spelled by school pupils. Fifty lists of 20 words each were constructed, and the words included in these lists were pronounced to the pupils of the various grades in the middle of the school year in the schools of 84 cities scattered throughout the United States. The data secured from these tests gave a total of 1,400,000 spellings by 70,000 school children. On the basis of these data, the 1000 words were divided into 26 groups according to difficulty. This will be understood by reference to the scale. (See scale inserted herewith.)

Group "A" consists of "me" and "do," and these words were spelled by 99% of the second grade pupils. At the other extreme, Group "Z," consisting of "judgment," "recommend," and "allege" were spelled by only 50% of the eighth grade pupils. The scale is simple, and easily understood. At the top of each column is shown the average per cent of the words spelled by each grade, except that report is not made upon any grade for per cents below 50. The blank spaces to the left, however, if filled in, would indicate in each case 100%, — that is to say, the eighth grade pupils spelled all of the words correctly from columns "A" to "N" inclusive.

**Giving a Test.**—A good test should be so difficult that no pupil in the grade will make a perfect score, and sufficiently

easy that most pupils in the grade will secure a fairly satisfactory score. In selecting words, therefore, to test the spelling ability of a particular grade, it would be well to choose the words spelled correctly by about 70% of the children of that grade. If pupils in the third grade were being tested, the best test would result from the use of words selected from column "L." A test, in order to be valid for individual pupils as well as for the group, should consist of at least 20 words. A smaller number of words would be equally valid for an entire school system, but the teacher will desire to know the standing of individual pupils, and so will need to use 20 words for the test. If 40 words were used, the results would be more reliable for individuals.

The tabulations of the scale are based upon tests given by the column method. This is the usual method of dictating words for pupils to spell by writing in columns. The Cleveland Survey shows that the returns from testing by this method differ very little from returns secured when the words are used in context. Other studies show that the contextual method (including words in complete sentences, the entire sentence being written) gives a slightly lower score. It is recommended, therefore, that teachers test by the column method. All that is necessary is that the pupils be given sufficient time to write a word before proceeding to the next word. The teacher should also be accommodating in re-pronouncing a word when necessary, in order to have it understood. Pronounce the words clearly, but do not sound them phonetically, or inflect them so as to aid the pupils in spelling. Give the meaning of words that sound like words with a different meaning and spelling. In case of difficulty in understanding a word, the best way to explain it is to use it in a simple sentence.

**Scoring the Papers.** — If there were 30 pupils in the third grade class above referred to, that would give a total of 600 spellings. Suppose that of these 600 spellings, 480 were



correct. Then 80% of the words were correctly spelled. Referring now to column "L" of the scale, it will be observed that the class, as a whole, is 7% above the standard of third grade pupils in the 84 cities which formed the basis for the scale. They are at the same time 8% below the standard for fourth grade pupils. Suppose that a particular child in the grade has spelled 17 words out of the 20, — that would mean a grade of 85%. This is better than the class average and only a little below the standard for the fourth grade. In the same way, the standing of each pupil in the grade may be determined.

In order to see at a glance the condition of her class, the teacher will find it worth while to arrange the scores for her grades in a distribution somewhat as follows:

TABLE 1. — DISTRIBUTED SPELLING SCORES FOR 30 THIRD GRADE PUPILS. STANDARD 73

Grade: <i>Third</i>		Date: _____													
Score . . .	40	45	50	55	60	65	70	75	80	85	90	95	100		
No. of pupils				1	1	2	4	5	7	6	3	1			

This table means that one pupil made a score of 55, one a score of 60, two a score of 65, four a score of 70, etc. This distribution emphasizes the needs of particular pupils. If the teacher of this particular third grade class can, by special work with the one pupil at 55, the one at 60, the two at 65, and the four at 70, bring these pupils up to the grade's standard, she will have a very satisfactory situation.

One of the advantages of the Ayres spelling scale is its simplicity and the ease with which it can be used. Because it contains the fundamental words of the language and the words on which the pupil should place his attention, the changes which it effects in the character of the spelling work



will be entirely in the right direction. To the extent that it does thus direct the attention to the proper kinds of words, we may expect that scores in particular cities will rapidly become higher than those indicated on the Ayres scale. This fact is indicated by the returns from the use of the Ayres scale in Boston, after considerable attention had been given by the teachers to the proper selection of word lists. Dr. Ayres himself has recognized this possible limitation, closing the discussion of his spelling scale with the following words :

“ In all such testing, it must be remembered that the present scale or any scale for measuring spelling attainment will become increasingly and rapidly less reliable for measuring purposes as the children become more accustomed to spelling these particular words. In proportion as these lists are used for the purposes of classroom drill, the scale will become untrustworthy as a measuring instrument. Probably the scale will have served its greatest usefulness in any locality when the school children have mastered these 1000 words so thoroughly that the scale has become quite useless as a measuring instrument.”

**Other Tests.**— While it is recommended that the grade teacher use only the Ayres scale in testing her pupils as individuals and her room for comparison with other rooms within the city or elsewhere, many teachers, and especially superintendents, will desire at least some information concerning other lists which have been used as spelling tests. The most notable of these are the Buckingham extension of the Ayres scale, the Iowa Spelling scale, the Buckingham scale, the Rice test, the Starch test, the Courtis Spelling test, the Boston Minimum list, and Jones' One Hundred Demons.

*Buckingham's Extension of the Ayres Scale.*— Dr. Buckingham's extension of the Ayres scale (first available in 1919) consists of the addition of 505 words chosen on the basis of agreements among spelling books. The words are added, for

the most part, to the upper end of the Ayres scale. This increases the number of words in the columns at the upper end of the scale and also extends the scale six steps to the right. The added words are not offered as constituting a fundamental vocabulary in the same sense as were the original 1000 words selected by Ayres. In using this extension, therefore, teachers should keep in mind that the added words have less value from the standpoint of social utility than the 1000 original words of the scale. The addition of these words, however, makes it possible to use the scale more extensively in upper grades and high school. It should be of particular value in testing the spelling efficiency of the pupils in the high school who are specializing in commercial studies.

*The Iowa Spelling Scale.* — This scale includes 2977 words from the written correspondence of Iowa people. Accuracy of each word was determined on the basis of 200 or more spellings by children in each grade. Thus, more than 650,000 spellings were used in each grade, or a total of nearly 4,750,000 in the seven grades. In all essential features the scale is an imitation of the Ayres scale. The placing of the words is determined in practically the same manner and the form of the scale is similar. It has decided value, however, as showing the possibility of basing the spelling work directly upon the words of a particular section of the country. The scale is published in three parts in order to reduce the error in the placement of words. Part 1 is a scale for grades 2, 3, and 4; part 2, a scale for grades 4, 5, and 6; and part 3, a scale for grades 6, 7, and 8. The large increase in the number of words makes the scale particularly valuable for individual testing.

*The Buckingham Scale.* — The work of Dr. Buckingham in evaluating a list of 50 words, has to date proved of value chiefly in calling attention to the importance of the proper selection of word lists, the difference in the difficulty of words,

and the methods to be used in the further study of words for spelling lists. The scale first appeared in 1913, and apparently has not come into general use in school testing and school survey work. The Ayres scale, which made its appearance a little later, is so convenient and so satisfactory that it has been extensively used by superintendents, bureaus of efficiency, and survey committees.

The fifty words resulting from the Buckingham study are given herewith, in the order of their difficulty. These words vary in difficulty by even distances, so that the scale, as it appears, is a step scale. Theoretically it should be used in such a way as to determine how far up the scale a pupil can spell successfully. It can be used in grades three to eight.

Dr. Buckingham, in deriving the scale, pronounced all of the words to the children in contextual form. In view of other studies which have been made, it appears that they could be used in column form with results slightly varying and equally satisfactory for comparative purposes. Although not in general use, the scale is mentioned because of the high quality of the scientific work involved in its formation. It has not been evaluated in terms of grade achievement. However, Dr. Buckingham is working on an extension of his scale. In time, he expects to extend it to include 1000 words and evaluate it in terms of grade achievement.

BUCKINGHAM'S FIFTY WORDS ARRANGED IN ORDER OF  
DIFFICULTY

- |            |             |               |
|------------|-------------|---------------|
| 1. only    | 9. pretty   | 17. cousin    |
| 2. even    | 10. nails   | 18. beautiful |
| 3. smoke   | 11. butcher | 19. touch     |
| 4. chicken | 12. Tuesday | 20. freeze    |
| 5. front   | 13. sure    | 21. forty     |
| 6. another | 14. answer  | 22. instead   |
| 7. lesson  | 15. nor     | 23. wear      |
| 8. bought  | 16. raise   | 24. tailor    |

- |             |               |               |
|-------------|---------------|---------------|
| 25. trying  | 34. against   | 43. telegram  |
| 26. minute  | 35. circus    | 44. saucer    |
| 27. pear    | 36. sword     | 45. saucy     |
| 28. towel   | 37. whistle   | 46. already   |
| 29. tobacco | 38. stopping  | 47. pigeons   |
| 30. whole   | 39. carriage  | 48. beginning |
| 31. button  | 40. guess     | 49. grease    |
| 32. janitor | 41. telephone | 50. too       |
| 33. quarrel | 42. choose    |               |

*The Rice Test.* — It was Dr. J. M. Rice, in his Forum articles of 1897, who first began the work of attempting a definite measurement of spelling. He gave three different tests, the number of children examined reaching nearly 33,000. The first test consisted of 50 words pronounced by the teachers for written spelling in the usual manner. The words used in this test were the following:

furniture	beggar	breakfast	Missouri
chandelier	plumber	chocolate	Alleghenies
curtain	superintendent	cabbage	independent
bureau	engine	dough	confectionery
bedstead	conductor	biscuit	different
ceiling	brakeman	celery	addition
cellar	baggage	vegetable	division
entrance	machinery	scholar	arithmetic
building	Tuesday	geography	decimal
tailor	Wednesday	strait	lead
doctor	Saturday	Chicago	steel
physician	February	Mississippi	pigeon
musician	autumn		

Dr. Rice had some question as to the value of word lists for spelling work, recognizing that spelling was useful only as a means for recording or communicating thoughts. This is the same point which we now recognize in different form; viz. that only the written vocabulary needs to be mastered for spelling purposes.

In line with this thought, Dr. Rice gave a second test to more than 13,000 children. This test contained 50 words placed in composition form. The following sentences were used, the underscored words forming the basis of the test:

"While running he slipped. I listened to his queer speech, but I did not believe any of it. The weather is changeable. His loud whistling frightened me. He is always changing his mind. His chain was loose. She was baking cake. I have a piece of it. Did you receive my letter? I heard the laughter in the distance. Why did you choose that strange picture? \*Because I thought I liked it. It is my purpose to learn. Did you lose your almanac? I gave it to my neighbor. \*I was writing in my language book. Some children are not careful enough. Was it necessary to keep me waiting so long? Do not disappoint me so often. I have covered the mixture. He is getting better. \*A feather is light. Do not deceive me. I am driving a new horse. \*Is the surface of your desk rough or smooth? The children were hopping. This is certainly true. I was very grateful for my elegant present. If we have patience we shall succeed. He met with a severe accident. Sometimes children are not sensible. You had no business to answer him. You are not sweeping properly. Your reading shows improvement. The ride was very fatiguing. I am very anxious to hear the news. I appreciate your kindness, I assure you. I cannot imagine a more peculiar character. I guarantee the book will meet with your approval. Intelligent persons learn by experience. The peach is delicious. I realize the importance of the occasion. Every rule has exceptions. He is thoroughly conscientious; therefore I do trust him. The elevator is ascending. Too much praise is not wholesome."

(The fourth and fifth year test ends with: "This is certainly true." The higher test includes all the sentences except the four marked with an asterisk.)

The third test given by Dr. Rice, and the one which he considered really more valid than any of the others, was a



composition test based upon a picture and a story told by the teacher. This test was valuable particularly in that it required pupils to choose their own words and to spell them.

The results on the third test are not tabulated by Dr. Rice, but we do have some tabulations on the first and second tests and the averages are given herewith.

TABLE 2. — RICE TESTS — 1895

GRADE	AVERAGE FIRST TEST (COLUMN LIST)	AVERAGE SECOND TEST (CONTEXT)
4	53.5	64.2
5	64.3	75.1
6	75.6	70.4
7	81.	78.8
8	84.2	84.4

Since Dr. Rice gave these tests to a sufficiently large number of pupils, the teacher may accept the averages given above as norms or standards of performance, and by comparison with them may determine the spelling ability of her own pupils.

The chief objection to the Rice list is that the words are not evaluated, and do not form a scientifically constructed scale. The words are given uniform values, but are far from being uniform in difficulty. In the Ayres and Buckingham scales, the words are assigned values according to difficulty. In the Rice test, a pupil gets as much credit for spelling an easy word as he does for spelling a difficult word.

*The Starch Test.* — Any one making use of the Starch test in spelling will do it with quite different purposes in mind than those for which he uses the Ayres scale. The words were secured by taking the first defined word on the even-numbered pages of the 1910 edition of the New International Dictionary. Proper names, technical words, and obsolete words were discarded from the list. The list, thus reduced to 600 words,

was arranged alphabetically according to the size of the words. These were then divided into six lists of 100 words each by assigning words in turn to the six lists. A test is made by using one of these six lists, which are assumed to be of equal difficulty as lists.

By using words selected at random from the entire English language, Starch proposes to test general spelling ability, and his tests will be found to be of service in the grammar and high school grades, provided the test is not permitted in turn to exercise an influence upon the teacher in determining the materials of the spelling lessons. The influence of the Starch test is surely in the direction of the old "spelling grind" described by Rice. The Starch lists contain such words as the following:

nunciature	conterminous	anthropometric
quarantinable	photosphere	imperturbation

Such words are manifestly not suitable for use with grade pupils.

*The Boston Minimum List.* — The Boston School Document No. 8, 1914, contains a minimum spelling list of 840 words. They are well selected, and similar in many respects to the Ayres list. However, they have not been evaluated for use as a standard test. The document containing this list, and a supplementary list of 2525 words, is no longer available except in libraries of departments of education. It is of interest chiefly in showing the tendency to get away from the old type of speller which contained 10,000 to 15,000 words, selected with little regard for use. The California list<sup>1</sup> is similar to the Boston list and is constructed along similar lines. It is of value for curriculum making in spelling, but not for testing.

*Jones' One Hundred Demons.* — Dr. Jones has given a list of the 100 words most often misspelled by pupils in written

<sup>1</sup> Bulletin No. 7, Chico State Normal, Chico, California.

work, as shown by his study involving the tabulation of 15,000,000 words. This list he has designated as the "spelling demons." The list has been widely used for testing, but to date it has not been sufficiently evaluated in terms of grade standards, although Dr. Jones promises such evaluation in the near future. The list appeals to children because of its simplicity, and its known difficulty. If a pupil thoroughly masters this list of "demons" he will very probably correct the spelling of most of the words which he has been misspelling. Dr. Jones did not find any pupil among the 1050 who missed as many as 100 words, 87 being the largest list for any one pupil.

The list of "spelling demons," together with their relative difficulty as shown by preliminary tests which Dr. Jones has summarized, follows herewith:

## FREQUENCY OF MISSPELLING OF THE JONES' 100 DEMONS

which 321	meant 247	minute 210	often 185
their 316	just 245	busy 209	writing 184
there 296	many 245	two 208	doctor 182
separate 283	too 243	much 206	very 182
hear 280	Tuesday 242	enough 206	though 181
here 278	knew 237	seems 205	among 179
said 275	lose 236	none 203	sure 179
been 273	week 235	does 203	tonight 174
says 273	can't 234	easy 202	forty 172
they 271	grammar 234	would 200	since 172
some 270	whole 231	whether 200	once 170
any 268	wear 230	loose 198	raise 169
Wednesday 266	every 228	could 196	trouble 168
done 263	instead 228	ready 196	choose 168
know 263	built 225	beginning 195	color 167
read ("red") 261	blue 224	heard 195	dear 166
piece 260	shoes 224	country 194	truly 166
don't 258	won't 221	business 194	early 166
break 257	wrote 220	ache 192	used 165
tear 255	cough 217	answer 191	friend 164



February 255	where 216	making 190	again 164
laid 252	write 216	always 188	hoarse 162
straight 251	buy 212	hour 187	guess 162
through 250	believe 212	tired 187	women 161
half 250	coming 212	sugar 185	having 158

*The Pupil's Own List of Misspelled Words.* — The final test of spelling is a gradual decrease in the pupil's own list of misspelled words. A necessary precaution in this connection is that pupils should not consciously avoid good words because they do not know how to spell them. They should be taught to use the dictionary instead of replacing good words by simpler words which they are able to spell. If every child is told to keep a list of his own misspelled words and to build up a spelling consciousness with the aid of the dictionary, and if he is urged constantly to extend his vocabulary and to study the choice of words in order to get appropriate and accurate expression, a pupil's spelling in regular written work may be considered as the best and the final test of spelling.

At stated intervals, a pupil should be encouraged to go over 8 or 10 pages of his written material and determine carefully the number of misspelled words. The teacher can help the child in doing this. But for the teacher to do it without the child's help has been in general the mistake of the past. In proportion as the number of misspelled words decreases, the child is improving in spelling.

While this test is not scientific, we can conceive of teachers making it even more valuable than scientific tests as they are frequently used. We do know that the time which a pupil spends upon his own list of misspelled words involves no lost effort; and that his spelling improves in the same proportion that this list is reduced. Indiscriminate drill in spelling, as indicated in the Butte, Montana, survey, must be replaced by attention to the needs of individual pupils. There were 278 of the Butte children, or over 18% of the total, who made scores of less than 60%, although the total

score for the city was 10.3% above the Ayres standard. Much time had been spent upon indiscriminate drill.

**The Practical Uses of a Spelling Scale.** — Teachers will find a spelling scale of very great use in their regular school work, aside from any supervisory use which the superintendent may make of the tests given. Tests administered under uniform conditions and with a scientifically constructed scale permit the teacher to compare one class with another very accurately. If the fourth grade teachers in a city system would agree among themselves to give a test on a certain day, they could then come together after the papers had been scored and find out, first of all, which room was doing the best work. This would be shown not only by the median score, but also by the total distribution which shows the number of pupils at lower as well as at higher levels.

After the teachers have agreed that a certain one of the fourth grade rooms has made, all told, the best score in the test, a second question naturally arises; namely, what method was used in securing these results with your children? This question suggests the second use which the teacher may make of the scale. She can test out different methods in her own room, or the particular group of fourth grade teachers to which we have referred may separate their rooms into groups of approximately equal ability and assign different methods for different groups. Then, at the close of a given period, — one, two, three, or six months, — they may again give a test and so determine which methods are most effective. If the teachers have been wise they have determined in great detail how the methods were to be applied and the amount of time to be devoted to the spelling work, so that the one thing which is upon trial is the *method of presenting the work*; such, for instance, as the column method, the contextual method, the method of studying at home or in the seat and then testing in class, the method of teaching in class with very little testing, and various other methods.

The above paragraph suggests a third point which teachers may try out by the use of a scientific scale; namely, the amount of time which can profitably be devoted to spelling. Dr. Rice, in his discussion of the spelling grind in 1897, showed that the time element had very little to do with results. We now know that this was because of the character of the spelling lists. When the words used in the spelling work with children are unintelligible to them, the results will be poor, regardless of the methods and the time devoted to the work. But if we assume words with correct social values, then the Ayres scale may properly be used for determining the amount of time which can be spent upon the spelling work with greatest profit.

A fourth use of the spelling scale has been suggested in asking the teacher to make a distribution of the grades. This use is to locate the spelling ability of individual children. By doing this, the teacher will probably find in her classes a small number of pupils who spell so well that it is unnecessary to require them to submit to any regular spelling drill. If such pupils are excused from spelling drill, being told merely to attend to their own misspelled words and to use the dictionary when in doubt, and if the teacher finds in future testing that these pupils do not lower their scores, then she may feel that she has saved their time for other more valuable work without detriment to them, so far as spelling is concerned. At the other end of the scale, however, will be pupils who spell very poorly, and it is only by use of the scale that these pupils can be located with any degree of accuracy. Taking these pupils as individuals, or as groups according to their several needs, the teacher can work in a definite manner, giving additional time to some pupils without boring others, and really follow out the injunction of William Hawley Smith to "put the oil where the squeak is." It is quite probable that this result of the use of the scale in spelling, as in writing, will in time become one of its most valuable contributions.

Some pupils will make low scores in their spelling work because of the lack of general intelligence; others, because of the lack of an adequate vocabulary, which can come only from reading; others because their attention has never been directed to the difficulties of words, etc., etc. The teacher will know that she is working at the problem in a definite manner, and that she is working only with the pupils who need attention. This she has known more or less before in a general way, but the use of a scientific scale permits her to know it beyond peradventure of a doubt.

It is not the purpose of the present work to discuss methods of spelling. The teacher is directed to other works dealing specifically with this problem.<sup>1</sup> The teacher will do well, however, to make her spelling work as specific as possible, both as to words and pupils. Many words spell themselves and require no attention, others are very difficult for large numbers of pupils. It is not only necessary to locate the words, but to analyze each word to see in what the difficulty consists. In short, drill which is general and blind must become specific and intelligent.

The discussion throughout has directed the attention of the teacher to the Ayres scale. Some teachers may properly ask if other scales may not at times be used to advantage. There will be no harm done in using other scales and the teacher may learn to use the Buckingham scale very effectively. The Jones' "Demons" have the advantage of being the words most frequently missed by school pupils. They are common words, and it is safe to assume that every pupil in the upper grades should study these words until he can spell the entire list without a mistake. In general, however, the Ayres scale is the one to use, for reasons which have been previously stated. There is this caution only, and that has

<sup>1</sup> Freeman, Frank N., "The Psychology of the Common Branches," Houghton Mifflin Company; Suzzallo, Henry, "The Teaching of Spelling," Houghton Mifflin Company; Cook and O'Shea, "The Child and His Spelling," Bobbs-Merrill Company, Indianapolis.

been anticipated by Ayres himself; namely, that as the scale is used more and more with the same pupils, a teacher should expect that gradually the scores will become higher. This, however, is quite satisfactory, since the words are of the right kind and since, by using the scale, the pupil's attention has been turned from unfamiliar, useless dictionary words to the words which he will use in his own work.

### BIBLIOGRAPHY

1. Ayres, Leonard P., "The Spelling Vocabularies of Personal and Business Letters," Division of Education, Russell Sage Foundation, New York City.
2. Ayres, Leonard P., "A Measuring Scale for Ability in Spelling," Division of Education, Russell Sage Foundation, New York City.
3. Buckingham, B. R., "Spelling Ability. Its Measurement and Distribution," Teachers College, Columbia University, New York City.
4. Des Moines Annual School Report, 1915. Section on "Spelling."
5. Jones, W. Franklin, "Concrete Examination of the Material of English Spelling," University of South Dakota, Vermillion, South Dakota.
6. Pryor, Hugh Clark, "A Suggested Minimal Spelling List," Chap. V, Part I, Sixteenth Yearbook of the National Society for the Study of Education.
7. Rice, J. M., "The Futility of the Spelling Grind," *Forum*, Vol. 23, pp. 163, 409.
8. Studley, C. K., and Ware, Allison, "Common Essentials in Spelling," Bulletin No. 7, State Normal School, Chico, California.
9. Wallin, J. E. W., "Spelling Efficiency in Relation to Age, Grade, and Sex, and the Question of Transfer," Warwick and York, Baltimore, Maryland.
10. The teacher or supervisor who is interested in the more intricate problems of establishing a spelling standard is referred to the following recent articles: Ballou, *School and Society*, 1917, Vol. 5, pp. 267-270; Ballou, *Educational Administration and Supervision*, 1915, Vol. 1, pp. 469-472; Kallom, *Educational Administration and Supervision*, 1917, Vol. 3, pp. 539-542.
11. Ashbaugh, Ernest J., "Iowa Spelling Scale," Extension Bulletin, Nos. 43, 54, and 55, University of Iowa.



## CHAPTER III

### THE MEASUREMENT OF HANDWRITING

THE writing supervisor had given Wilbur a grade of 95. Wilbur was dissatisfied. When the supervisor next came to the building, Wilbur made known his dissatisfaction, and asked why his grade was not higher. The supervisor answered that 95% was a good grade, that she never gave 100%, and that there was opportunity for him to further improve his work. Wilbur answered that he had received 95% from the fourth grade up, and he knew that he was writing much better than in any previous grade. The supervisor had no conclusive or satisfactory argument. She resorted to her authority as teacher, and left Wilbur still dissatisfied. What teacher has not had a similar experience with reference to the grade in writing?

This situation is rapidly changing in the public schools. Writing can be definitely measured, and the ratings can be made so accurately that the pupils themselves fully understand and appreciate that exact justice has been done. This has been brought about by the development of scales for the measurement of handwriting.

If a teacher has not been accustomed to make use of scales and standardized tests in her work of grading, she would do well to begin with the subject of writing. Writing is one of the mechanical subjects and one of the most easily and quickly measured. In order to avoid confusion on her part, she should study and practice scientific measurement in this subject alone until she has become reasonably proficient. It will be well for the teacher to read through a large number of the works mentioned in the bibliography at the close

of this chapter, and as a beginning in this work, particular attention is called to numbers 1 and 2.

The first scale in handwriting was developed by Dr. E. L. Thorndike, of Teachers College. It is based upon general merit in handwriting as determined by the judgment of a large number of competent graders. Thorndike's scale is widely used at the present time, and many think that it gives more satisfactory results than any other. It had, originally, the disadvantage<sup>1</sup> of being mechanically inconvenient, and for that reason the Ayres scale has become much more widely used.

The Ayres scale consists of twenty-four samples of writing, eight each of vertical, semi-slant, and full slant style. The scale is arranged on a heavy sheet of paper 9" high and 36" wide, in the form of the following diagram:

	20	30	40	50	60	70	80	90
A .								
B .								
C .								

It is so convenient in form that it may be placed in the schoolroom, where pupils may compare their handwriting with it at any time. This is desirable, and it is recommended that every schoolroom in which there are intermediate and upper grade pupils should have a copy of the Ayres scale available for pupils as well as for teachers. (See pp. 28-35.)

**What to Measure.** — Ordinarily the teacher will measure only two elements in handwriting; namely, speed and quality. By speed is meant the number of letters written per minute. By quality is meant general merit, or what the teacher indicates when she gives a grade in writing. Speed is determined

<sup>1</sup> A defect since remedied in large measure.

by simply counting the number of letters written during a given time and reducing to the one-minute basis. It is quality or general merit which is measured by the use of the writing scale. These terms are relatively simple, and their significance will appear during the further discussion. It is just as well for the teacher to begin by giving a regular test, and in this manner to apply herself to the work of mastering the details of grading and evaluating papers in handwriting.

**Giving the Test.** — In order to make the test valid for comparative purposes, uniform conditions must prevail. The rules of the game are simple, and the teacher should follow them carefully, since it is only in this way that valuable comparison will be made possible. The directions for tests in handwriting are so generally standardized at the present time that comparison is possible, not only within the class, but one room with another and even one school system with another. The invariable aim is to secure results in such form as to make them easily comparable with like results obtained elsewhere. The rules are as follows:

1. The copy must be simple enough for second grade pupils. While it is not necessary to use the same copy each time, it should be similar in difficulty. A copy which has been much used is the line: "Mary had a little lamb." Others have used the entire first stanza of this selection. Another copy which has been used is "Sing a song of sixpence, a pocket full of rye." The idea is to have a simple, easily understood copy, which will not deter the pupil in his speed test. Some tests have been given with copy which was too difficult, making the results in speed unsatisfactory for comparative purposes.

2. Before the test is given, the copy should be memorized by all of the pupils. The purpose of the test is to determine speed and quality of handwriting. If the pupil must stop and think, he falls behind in speed. In one survey a rather difficult copy was placed in the hands of the pupils. They



were instructed to write the copy, repeating the same during the period of the test. The results were so unsatisfactory that speed was not reported upon by the survey committee. In addition to having the copy committed, it is a good plan to place the same upon the blackboard at several different places, so that any pupil who does happen to forget for a moment may reassure himself by a glance at the copy.

3. The time for the test should be exactly two minutes. In order to make sure that all pupils start together, it is well to rehearse the details before actually starting the test. This makes sure that all pupils understand, clears away any confusion, and so secures the test papers in reliable form.

4. Everything should be in readiness for the test before the pupils begin. This means that every pupil must have paper, a good pen, ink, and the copy committed. In order to make sure that all have pens, it is well to ask every pupil in the room to hold up the pen (or pencil, if used in second or third grade). Since the teacher will want to use the results of the test for the benefit of individual pupils, it is well at this point to place certain items at the head of the paper. The usual items are — name, grade, building, city, and date. If for any reason it is desired to make the test impersonal, these items may be omitted, or placed on a separate card with a number scheme as a key.

5. When all is ready, the teacher gives some simple directions. "Write as well as you can at your usual speed, using the following copy: 'Mary had a little lamb.' Write the copy again and again until I say 'stop.' At the command, stop at once, even if in the middle of a letter." After this explanation has been given the teacher says, "All in position. Dip the pens. Pens up. Begin."

6. In exactly two minutes, pupils should be given the order to stop, and required to place their pens on the desk.

7. At this point the teacher may save herself considerable work by having the pupils count the number of letters in the

copy. It is suggested that pupils place this number below the copy to the right, using pencil for the same, and then divide the number by two, thus reducing the score to a one-minute basis, as  $2)146$ . The papers may then be collected

73

in the usual manner.

**Scoring for Speed.** — The speed is calculated in terms of the number of letters written per minute. The test is given over a two-minute period in order to reduce the error. Some examiners have used other units, as three or four minutes, but evidence is not at hand that the results have been improved. In the first report upon speed in handwriting,<sup>1</sup> two minutes was made the basis of the test, and this unit has quite generally been used in later tests. The practice is common, also, of reducing to the one-minute basis, thus making comparison easy.

The speed measurement is secured by counting the letters in the pupil's copy and dividing by two. Although the pupils have been asked to count the number of letters, the teacher should carefully check the results. The teacher may reduce her work by knowing the total number of letters in the copy used, multiplying by the number of repetitions of the full copy, then adding the extra letters. Suppose a particular pupil has written the copy, "Mary had a little lamb," eight times, and has written the first three words the ninth time. The teacher in figuring the number of letters will multiply 18 by 8 which gives her 144 and then add the number of letters in the three words, — "Mary had a," namely, eight. This gives a total of 152 letters. Dividing by 2 she gets the pupil's score, 76 letters per minute. In case the teacher gets a result different from the pupil's result, the same should be placed in the lower right-hand corner, the pupil's figure being crossed out. This completes the scoring of the papers for speed.

<sup>1</sup> Wilson, G. M., "The Handwriting of School Children," *Elementary School Teacher*, 11: pp. 540-543. This is the first known attempt to fix a standard for speed in handwriting.

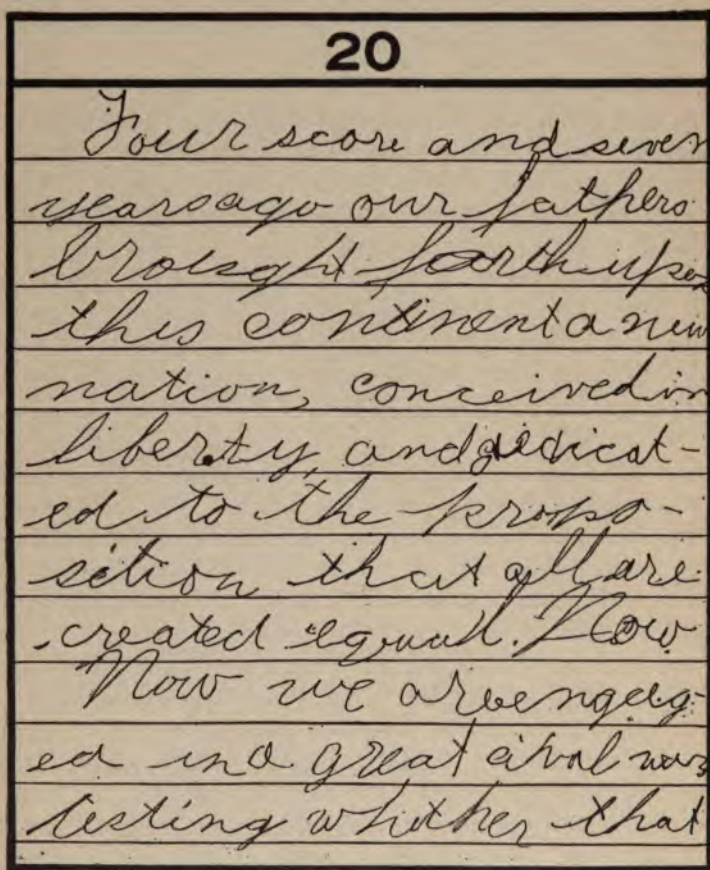


FIG. 1. — Ayres handwriting scale (pp. 28-35). The copy shown herewith is the so-called Gettysburg edition.

## 30

Four score and seven years ago our fathers brought forth upon this continent a new nation, conceived in liberty and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war testing whether that nation or any nation so conceived and so dedicated can



## 40

Four score and seven  
years ago our fathers  
brought forth upon  
this continent a new  
nation, conceived in  
liberty, and dedicated  
to the proposition  
that all men are  
created equal.

Now we are engaged  
in a great civil war,  
testing whether that

## 50

Four scores and seven  
years ago our fathers  
brought forth upon this  
continent a new nation,  
conceived in liberty,  
and dedicated to the  
proposition that all  
men are equal. Now  
we are engaged in a  
great civil war, testing  
whether that nation, or  
any nation so conceived

60

Four score and seven years ago our fathers brought forth upon this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal

Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated

## 70

Four score and seven  
years ago our fathers  
brought forth a new na-  
tion, conceived in liberty,  
and dedicated to the  
proposition that all men  
are created equal.

Now we are engaged  
in a great civil war,  
testing whether this  
nation, or any nation  
so conceived and so ded-



**80**

Fourscore and seven  
years ago our fa-  
thers brought forth  
on this continent, a  
new nation, conceiv-  
ed in Liberty and  
dedicated to the prop-  
osition that all men  
are created equal.

Now we are en-  
gaged in a great  
civil war, testing

90

Fourscore and seven  
years ago our fa-  
thers brought forth  
upon this continent  
a new nation, con-  
ceived in liberty,  
and dedicated to  
the proposition that  
all men are created  
equal. Now we are  
engaged in a great  
civil war testing

**Scoring for Quality.** — The teacher will be surprised how quickly she can learn to grade papers by using the Ayres scale. While it is helpful to have a demonstration and some practice in a teachers' meeting, this is not at all necessary, and the teacher who is patient and willing can train herself very quickly to use this scale and to secure satisfactory results. The teacher should give herself preliminary drill of at least an hour or two. If this drill is divided into half hour periods, and continued during a considerable part of a week, the teacher will become reasonably uniform in grading papers, and will feel competent to score the papers from the test in her room. At this point it would be well for her to consult an expert, in case one is available. This expert by a little observing and advising will correct any marked defect, — such as a uniform tendency to grade too low or too high. In the absence, however, of a teacher, a supervisor, or a superintendent, in the system, who can give this expert help, a teacher need not be deterred. She can master the details, working entirely alone.

Directions for grading a sample, while not uniform, have in mind the common object of helping the teacher to locate the specimen on the scale which most nearly corresponds in merit with the pupil's copy. Apparently the best way to do this is to glide the pupil's copy back and forth underneath the scale, comparing it with one sample after another in the scale until a decision is reached as to which sample most nearly corresponds with the pupil's copy. The teacher will frequently have difficulty, and especially where the pupil's copy is better, for example, than 50 on the Ayres scale, but not as good as 60. Some scorers recommend the use of intermediate units in such cases, permitting the teacher thus to indicate 54, 56, or whatever the proper value may appear to be. Practice on this point varies. If the number of papers to be scored is not too large, intermediate values may be used.

The score for quality when determined upon should be placed in the upper right-hand corner of the paper.

**Recording the Scores.**—From the beginning the teacher should acquire the habit of distributing her scores, showing both speed and quality on a single sheet. This will be found exceedingly helpful. Table 3, which follows herewith, shows such a distribution for a sixth grade. By reference to this, it will be seen that of the 33 pupils in the grade, 2 are writing at quality 20 (see totals at the bottom of the sheet), 4 at quality 30, 5 at quality 40, 8 at quality 50, 8 at quality 60, 5 at quality 70, and 1 at quality 80. The middle<sup>1</sup> score on the basis of quality will fall therefore in the group of 8 at 50 and this is noted below as the median quality.

TABLE 3. — DISTRIBUTION OF SCORES FOR A SIXTH GRADE

	20	30	40	50	60	70	80	90	TOTALS FOR SPEED
1- 20 . . . .									
21- 30 . . . .	I	I							2
31- 40 . . . .		I	I	I		I			4
41- 50 . . . .	I	I	2	2	I				7
51- 60 . . . .		I	I	2	3	I			8
61- 70 . . . .			I	2	2	2			7
71- 80 . . . .				I	I		I		3
81- 90 . . . .						I			I
91-100 . . . .					I				I
101-120 . . . .									
121-140 . . . .									
141-160 . . . .									
161-180 . . . .									
181-200 . . . .									
Totals for Quality	2	4	5	8	8	5	I		33

Median Quality — 50

Median Speed — 56

The totals for speed are indicated in the right-hand column. It is observed that the median speed falls between 51 and 60. In this particular case, however, the teacher has determined

<sup>1</sup> See explanation of middle score, or median, p. 261. Since there are 33 papers, the middle score in this case will be that of the 17th paper from either end.

the exact median for speed, and it is recorded below as 56. To determine the exact median for speed all that is necessary is to arrange the papers in order, from lowest to highest on the basis of speed, then count in to the middle paper. In this particular case the middle paper would be the 17th one from either end, and it appears that the 17th one had a speed of 56 letters per minute.

**Standard Scores.** — With the scores fully tabled the teacher's next question naturally is, "How does the writing of my pupils compare with others, and what are the standards?" She wonders if sixth grade pupils should show a range in quality from 20 to 80, and if a median quality of 50 is too low. In speed she notes that they are distributed from less than 30 to nearly 100. This means that some of the pupils are writing three times as rapidly as others. How rapidly should they write? So far as known this question was first raised only six years ago, and at that time a tentative standard for speed was indicated on the basis of results from a single city system.

TABLE 4. — STANDARDS OF SPEED<sup>1</sup>

GRADES	1	2	3	4	5	6	7	8
1. Cleveland . . . . .					60	70	76	80
2. Kansas City (May, 1915) .			53	64	69	76	76.5	
3. Denver Survey . . . . .			36	50	54	63	66	69
4. South Bend (May) . . . .		33	48	63	77	82	93	105
5. Freeman's 56 cities . . . .		31	44	51	59	63	68	73
6. Brookline . . . . .					76	87	90	98
7. Newton . . . . .					73	85	94	102
8. Missouri Training Schools .					80	92	92	102
9. Iowa, 33,569 children . . .	29	39	50	62	65	73	75	76

Now, however, it is possible to indicate a standard based upon results obtained from all parts of the country, and to

<sup>1</sup> Decimals largely omitted.



indicate rather definitely how well pupils in any particular grade should write.

Table 4, given herewith, shows the median attainment in speed for Cleveland, Kansas City, Denver, South Bend, fifty-six cities combined, Brookline, Newton, the Missouri Training Schools, and over 33,000 Iowa children.

From this table it will be seen that sixth grade children from different parts of the country are averaging from 63 up to 92 letters per minute. It should be noted, however, that the 82 for South Bend is a May average and was secured by special attention after a test given earlier in the year had shown the need for improvement. It is apparent, then, that the particular sixth grade shown in Table 3 is quite definitely below standard, if we take as a standard the performance of other sixth grade children throughout the country. In this connection, it may be well to note two proposed standards made by men who have given considerable thought and attention to the subject.

TABLE 5. — STANDARDS FOR SPEED IN HANDWRITING

GRADES	2	3	4	5	6	7	8
Freeman . . . . .	36	48	56	65	72	80	90
Starch . . . . .	31	38	47	57	65	75	83

Tables 4 and 5 will give plenty of opportunity for comparison with actual performance and with proposed standards, to enable the teacher to judge of the writing in her own room. It appears that the median speed of 56 for her sixth grade is lower than the sixth grade median of any system appearing in Table 4, and indicates that the teacher should increase the speed of writing in this particular grade. She should at least aim to reach 63, the average of Freeman's 56 cities, the average also for Denver and the lowest sixth grade median appearing in Tables 4 or 5.



**Standards for Quality.** — In measuring quality for comparative purposes it is necessary to use one of the standard scales of handwriting. Not all studies in the measurement of handwriting have made use of the Ayres scale, but Table 6, given herewith, shows several returns in the Ayres scale and will permit comparison.

TABLE 6. — QUALITY IN HANDWRITING (Ayres)

GRADES	1	2	3	4	5	6	7	8
Brookline . . . . .					44	46	47	49
Cleveland . . . . .					45	48	50	55
Denver . . . . .			26	31	38	43	51	57
Newton . . . . .					48	51	50	53
South Bend (May) . . . . .		45	49	49	49	53	56	54
Missouri Training Schools . . . . .					41	42	45	47
Iowa median . . . . .	28	36	40	44	49	52	57	61
Freeman, 56 cities . . . . .		44	47	50	55	59	64	70

It will be observed from this table that quality in handwriting for the sixth grade has ranged from 42 in the Missouri Training Schools to 59 in the 56 cities reported by Freeman. It appears therefore that the particular sixth grade reported in Table 3, is writing better than the sixth grade pupils in the Missouri Training Schools, Brookline, Cleveland, and Denver, but not so well as those in South Bend, Newton, Iowa, or Freeman's 56 cities.

The standards proposed by Freeman and Starch for quality are likewise given herewith :

TABLE 7. — STANDARDS OF QUALITY IN HANDWRITING

GRADES	2	3	4	5	6	7	8
Freeman . . . . .	44	47	50	55	59	64	70
Starch . . . . .	27	33	37	43	47	53	57

It will be observed that the particular sixth grade writes better than the standard indicated by Starch, but not so well as the standard indicated by Freeman.

**Social Standard of Writing.** — In attempting to set up standards, there is one danger which school people are likely to encounter, and that is the danger of considering writing as a school exercise, wholly apart from the social and business demands of life outside the school. In the last analysis it is this latter which should determine the proper standards. While it is difficult to get at the standards required by society, there are at least some evidences of social standards of handwriting. Dr. Ayres has constructed a special handwriting scale for the Municipal Civil Service Commission of New York City. On the basis of this scale, the Commission considers that applicants pass in handwriting if they make a grade corresponding to quality 40 of the Ayres public school scale. Where handwriting is a special requirement a grade equal to quality 50 is required. These standards are lower than the Freeman standard for the sixth grade, and correspond fairly well with the Starch standard. However, sixth grade pupils will be in school two years longer, and under the present régime will write and continue to improve their writing for two years. This naturally raises the question as to whether the school standard for handwriting is not an artificial one, whereas it should be based directly upon the demands of society.

There is additional evidence on this matter, as reported on page 24 of the First Iowa Elimination Report, as follows: "One hundred graduate students of Teachers College wrote at a median quality less than 50. Three hundred Indiana teachers in Perry, Green, and Ripley Counties wrote at median qualities less than 50. One hundred inquiries for help received by the Social Service Bureau of New York City showed a median quality less than 50. One hundred applications for positions ranging from \$10 a week to \$5000 a year, received by the

Social Service Bureau of New York City, showed a median quality of 60. Signatures on 100 bank checks showed a median quality of 41. 256 signatures on a hotel register showed a median quality of 41.1." It appears from the above that the adult social standard is fully satisfied by a quality of 50 for practically all purposes. Even in the case of applicants for positions, where there is a special incentive for good writing, the median rises only to 60. On the basis of social usage, therefore, it appears that a quality of 60 on the Ayres scale should be accepted as satisfactory for any grade of school work, and that when pupils have attained a quality of 60, with reasonable speed, they should be excused from further writing drill unless a pupil voluntarily chooses to continue. It will be observed from Table 6 that most 7th and 8th grade medians fall between 50 and 60. A quality of 60 therefore appears reasonable and attainable for upper grades. A higher standard except for special commercial positions would be artificial and unreasonable.

What should be accepted as a reasonable *speed* from the standpoint of society has not been determined in any authoritative manner. It is quite probable that a speed of 60 or 70 letters per minute is sufficient to meet almost any situation. It would seem, therefore, that a teacher who brings her pupils to a *quality of 60 and a speed of 60* has prepared them to meet the handwriting demands of society. Many pupils, because of special interests or superior abilities, will prefer to go above this, easily meeting the extreme social demands where handwriting of superior quality is required.

**Remedial Instruction.** — When the sixth grade teacher has distributed her scores as shown in Table 3, and has decided what should be considered a reasonable standard in speed and quality for sixth grade pupils, her next question is how to remedy the situation for the pupils who are below standard in speed and quality. Studies have indicated that merely extending the time for the writing work will not solve the

problem. In fact, there is much evidence that children write too much and fall into careless habits for that reason. The story of how to remedy the defects is a long one, and will not be taken up fully in this discussion. The teacher is referred to other sources, particularly to the "Teaching of Handwriting," by Frank N. Freeman. There are certain phases of the work of remedying defects, however, which have been subjected to definite measurement.

Freeman has constructed a series of writing scales or charts, based upon the most common defects of the pupils' writing. These scales or charts deal respectively with — 1, Uniformity of slant; 2, Uniformity of alignment; 3, Quality of line; 4, Letter formation; 5, Spacing. Each chart contains three qualities of excellence, illustrating good, average, and poor qualities of handwriting from the standpoint of the characteristic dealt with in the particular chart.

The teacher who is especially interested in writing, and especially the writing supervisor, will find it worth while to make use of Freeman's analytical charts. By carefully selecting samples of the pupils' writing she can for her own use make up charts similar to the Freeman charts, thus having available for showing to the pupils samples that illustrate desirable and undesirable features under uniformity of slant, uniformity of alignment, etc.

Table 8, given herewith, should prove especially helpful, as it indicates the causes for the various defects. The teacher and pupil should work together in applying this table to the pupil's writing. If a pupil is writing with too much slant, the teacher will do well to study the pupil in the light of the five suggested causes. It may be a matter so simple as having the paper in the wrong position — and so with other defects. It is a matter of studying the situation with the particular pupil, analyzing the defect, finding the cause, and helping the pupil to apply the remedy.

TABLE 8.—ANALYSIS OF DEFECTS IN WRITING AND THEIR CAUSES<sup>1</sup>

DEFECT	CAUSES
1. Too much slant . . . .	(1) Writing arm too near body. (2) Thumb too stiff. (3) Point of nib too far from fingers. (4) Paper in wrong position. (5) Stroke in wrong direction.
2. Writing too straight . . .	(1) Arm too far from body. (2) Fingers too near nib. (3) Index finger alone guiding pen. (4) Incorrect position of paper.
3. Writing too heavy . . . .	(1) Index finger pressing too heavily. (2) Using wrong pen. (3) Penholder too small diameter.
4. Writing too light . . . .	(1) Pen held too obliquely or too straight. (2) Eyelet of pen turned side. (3) Penholder too large diameter.
5. Writing too angular . . .	(1) Thumb too stiff. (2) Penholder too lightly held. (3) Movement too slow.
6. Writing too irregular . .	(1) Lack of freedom of movement. (2) Movement of hand too slow. (3) Pen gripping. (4) Incorrect or uncomfortable position.
7. Spacing too wide . . . .	(1) Pen progresses too fast to right. (2) Too much lateral movement.

The teacher may find it advisable to extend the list of defects, and this can doubtless best be done by making use of the analytical score card for handwriting, developed by Dr. C. Truman Gray of the University of Texas. It is indicated herewith, Figure 2. Dr. Gray's score card is in many respects more complete than the detail of defects listed by Dr. Freeman.

The teacher will do well to enlist the pupil fully in the attempt to improve his writing. For the most part the pupil

<sup>1</sup> F. N. Freeman's "The Teaching of Handwriting" in the *Riverside Educational Monographs*, page 72, published by Houghton, Mifflin Company. By special permission of the publishers.

**FIGURE 2. — STANDARD SCORE CARD FOR JUDGING HANDWRITING**

(Devised by C. Truman Gray)

Pupil. . . . . Age. . . . . Date. . . . .  
Grade. . . . . School. . . . .  
Teacher. . . . .

[illegible]



simply knows that his writing is poor. He doesn't know why it is poor, and he is given no help in applying proper remedies. If he realizes, for instance, that it is a question of slant, or of uniformity in spacing, or uniformity in height, or neatness, — that is, if he can be made to place his attention upon some particular defect and work toward the correction of that defect, he can feel that he is working toward some definite end and not merely drilling aimlessly upon writing. The teacher's business here is to teach, not to scold, not to find fault. The teacher may not find it advisable to use the Gray score card, so far as actually scoring the pupils' work is concerned, but she can use it along with Freeman's suggestions in discovering with the pupil the defects which need remedying. In time the teacher may be able to construct a chart showing letter defects similar to Freeman's, but made up entirely from work of her own pupils. Freeman's chart<sup>1</sup> shows the correct form of a letter, together with the usual defects. It will help to furnish an answer to the pupil's "Why," when he asks why he was marked down in writing. All pupils appreciate being treated with consideration and given an opportunity of doing a reasonable amount of thinking in connection with their work.

**Locating the Individual.** — The discussion under remedial instruction shows the necessity of locating the individual. It is suggested that the teacher be not satisfied with the distribution as indicated in Table 3, but go a step farther, placing in the names of the particular pupils, as in Table 9. This will individualize the work, and will also make it more intelligible to the children. Raising the score in quality for her room then becomes a question not of blind unintelligent drill, but a question of improving the work of John, Mary, Jane, William, etc. In fact, taking the particular sixth grade as an example, and accepting quality 60 as the standard, it is observed that 14 of the pupils are already writing satisfactorily.

<sup>1</sup> The Teaching of Handwriting, page 135.

From the standpoint of speed, 12 are writing above 60 and it is possible that some of the 8 writing between 51 and 60 are on a satisfactory basis. This analysis of the situation limits the teacher's efforts to particular pupils, and enables her to apply her instruction where it is most needed. It also eliminates useless drill. At least two of the pupils writing at quality 60 or above are below in speed. These are Jeanette and Mark. Four others, Grace, Lily, Henry, and David, are also below in speed or just on the line. Four who are satisfactory in speed are below in quality. These are Bruce, Ruth, Bert, and Thomas. The eight to the right and below the heavy lines are satisfactory in speed and quality, and further drill by them may be left to choice. If this plan were generally followed in school systems, a large amount of effort would be released in handwriting alone, for application along other needed lines.

TABLE 9. — DISTRIBUTION OF SCORES FOR A SIXTH GRADE

	QUALITY 20	30	40	50	60	70	80	90	TOTAL FOR SPEED
(Speed)									
1- 20 . .	John	Mary							2
21- 30 . .		Jane	Orie	Kate		Mark			4
31- 40 . .		Luther	Sarah	Carrie	Jeanette				7
41- 50 . .	William	Epsie	Hazel						
51- 60 . .		Bertha	Joe	Grace	David				8
		Wilber	Paul	Lily	Henry				
61- 70 . .			Bruce	Ruth	Eldon	Bess			7
				Bert	Ina	Frank			
71- 80 . .				Thomas	Mildred		Doris		3
81- 90 . .						Helen			1
91-100 . .					Jacob				1
101-120 . .									
Totals for Quality . .	2	4	5	8	8	5	1		33

**Proportion of Children at Standard Quality.** — Figure 3, given herewith, shows a distribution of upper grade pupils in Cleveland, Ohio. Computation shows that 3303 of the

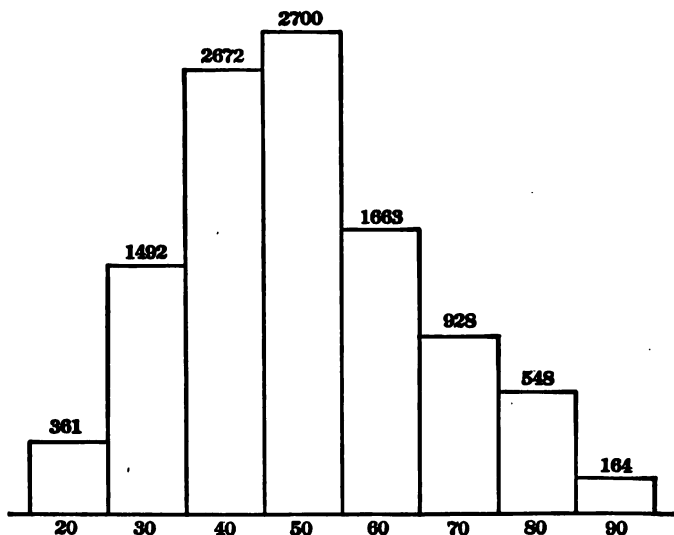


FIG. 3. — Number of pupils writing at each quality from 20 to 90. Data from 10,528 pupils in four upper grades (Cleveland Survey, p. 70, "Measuring the Work of the Public Schools"). 31.3 % at 60 or above.

children, or a total of 31.3%, were writing at quality 60 or above. The Springfield, Illinois, survey showed that 33.3% of the upper elementary grade pupils were writing at 60 or above. In the Butte, Montana, survey 23.8% of the pupils in grades 2 to 8 were writing at quality 60 or above. In Kansas City, in 1915, 16.4% of all pupils were writing at quality 60 or above. In the three upper grades in Kansas City these percentages were as follows:

- Fifth grade — 25.1% at quality 60 or above.
- Sixth grade — 39.7% at quality 60 or above.
- Seventh grade — 48.4% at quality 60 or above.

This means that in the seventh grade in the Kansas City schools, practically half of the children were writing at a satisfactory standard of quality, and should have been excused from further drill.

These figures taken from city reports and surveys make it evident that many upper grade pupils should properly be excused from further writing drill, and that our illustrative sixth grade throughout this chapter is quite representative in its distribution of writing ability in an intermediate or upper

			79		
			78		
			78		
		68	77		
		66	77		
		65	75	89	
		62	72	87	
	55	62	71	84	
	55	61	71	83	
	52	60	71	83	
	50	60	71	81	
47	50	60	70	81	93

FIG. 4. — Speed records of 36 sixth grades, Cleveland.

grade. The procedure recommended for this grade should be applied quite generally. Some teachers, however, may want to require pupils to *reach and maintain in all written work* a quality somewhere above 60, even as high as 70, before excusing them from further drill in writing. Some

pupils who are excused from drill may prefer to continue until a higher standard is reached.

**The Writing of an Entire School System.**—Above in Figure 3, the writing scores for an entire school system above the third grade are thrown together into a single distribution. There are various ways in which these data for a school system may be used to advantage, the following being particularly useful:

1. A grade in one building or part of the city may be compared with the same grade in other buildings or parts of the city. Figures 4 and 5 show this detail in median

			49			
			49			
			49			
			49			
			48	54		
			48	54		
			47	54		
		44	47	53		
	39	43	46	53		
	38	43	46	51	58	61
	38	43	46	51	58	61
30	35	41	45	50	56	60

FIG. 5.—Quality records of 36 sixth grades, Cleveland.

speed and quality for the sixth grade of the Cleveland schools. Any teacher may locate her particular grade in these distributions, and so see its rank in terms of the median scores.

2. A city may be compared with another city or with an established norm or standard. Table 10 will aid in this work.

It is valuable as a means of showing quickly and forcefully the relative standing of the city in question. Here Cleveland is compared with 12 other cities in speed and quality, as follows:

TABLE 10. — SPEED AND QUALITY IN HANDWRITING, CLEVELAND, OHIO

	AVERAGE SPEED			
	5th grade	6th grade	7th grade	8th grade
12 other cities . . .	57	65	75	83
Cleveland . . .	62	69	73	78

	AVERAGE QUALITY			
	5th grade	6th grade	7th grade	8th grade
12 other cities . . .	43	47	53	57
Cleveland . . .	45	48	50	55

The Freeman standards have been much used for such comparisons.

3. After building scores and medians have been ascertained, a superintendent of a school system may desire a total city summary. Table 11, following herewith, permits comparison of the writing in any particular building with writing in the other buildings and the city average. This will be particularly interesting and stimulating to teachers, principals, and supervisors.

In speed, Cleveland excels the other cities in grades five and six, but is below in grades seven and eight. However, there is some evidence that Cleveland is more nearly right in the matter of speed in writing than is the average of the twelve other cities. Likewise in quality, grades five and six of the Cleveland schools do better than the average of the twelve other cities, but grades seven and eight are below the average of the twelve cities.



TABLE 11.<sup>1</sup>—DISTRIBUTION OF MEDIAN SCORES IN QUALITY OF PENMANSHIP BY SCHOOLS AND GRADES. (Salt Lake City)<sup>2</sup>

GRADE	III	IV	V	VI	VII	VIII
Emerson School . . . . .	9.6	9.5	12.5	10.9	12.4	11.3
Forest School . . . . .	9.3	10.4	10.2	9.9	11.9	13.2
Grant School . . . . .	8.2	10.1	10.9	10.9	10.4	
Hamilton School . . . . .	11.9	10.1	11.5	12.9	12.5	
Jackson School . . . . .	10.7	10.7	9.9	10.5	11.4	13.
Jefferson School . . . . .		9.5	11.3	11.5	11.3	11.6
Lafayette School . . . . .	10.5	11.3	10.6	10.3	12.2	14.7
Lincoln School . . . . .	9.0	9.2	9.0	11.	11.2	
Lowell School . . . . .	8.6	10.6	11.7	11.8	14.	14.6
Onequa School . . . . .	10.5	11.6	10.9	9.9	12.2	13.5
Oquirrh School . . . . .	8.7	10.7	12.2	13.3	12.1	
Poplar Grove School . . . .		9.5	9.8	11.3	11.6	12.4
Riverside School . . . . .	9.4	12.7	9.8	11.	12.	12.2
Summer School . . . . .	10.2	13.8	12.4	12.2	12.7	13.9
Training School . . . . .	7.1	9.0	9.8	9.6	11.6	12.5
Wasatch School . . . . .		12.7	13.4	11.3	12.4	12.3
Washington School . . . . .	8.9	9.7	9.5	10.7	11.2	
Webster School . . . . .	7.6	11.1	10.7	12.1	12.8	11.6
Whittier School . . . . .	9.1	11.7	11.4	12.0	12.8	14.7
For the City . . . . .	9.2	10.7	11.0	11.3	12.2	12.8

Table 12 shows a total distribution of quality in writing for Salt Lake City. It will be observed that this table gives a different view from the distribution of medians shown for Cleveland in Table 6. It gives a worth while bird's-eye view of the writing for the entire city. This distribution is particularly valuable to the superintendent and supervisors in showing the work yet needed on handwriting. Quality 12 of the Thorndike scale corresponds to 60 of the Ayres scale.

<sup>1</sup> The scores in Tables 11 and 12 are in terms of the Thorndike scale to be explained further on in the chapter.

<sup>2</sup> *Salt Lake City Survey*, page 148.

TABLE 12.—THE DISTRIBUTION OF SCORES IN QUALITY ON 3685 SAMPLES OF PENMANSHIP BY GRADES. (Salt Lake City) <sup>1</sup>

SCORE	III	IV	V	VI	VII	VIII
4 . . . . .	3					
5 . . . . .	4					
6 . . . . .	21	5		3		
7 . . . . .	55	30	3	3	2	
8 . . . . .	85	63	59	26	8	
9 . . . . .	196	175	147	117	70	28
10 . . . . .	46	37	23	38	12	4
11 . . . . .	102	152	190	53	163	97
12 . . . . .	44	60	65	92	91	81
13 . . . . .	39	101	98	87	189	84
14 . . . . .	11	38	41	52	68	50
15 . . . . .	4	12	15	20	31	35
16 . . . . .	4	9	4	10	24	61
17 . . . . .		4	1		2	10
18 . . . . .		1		1	2	22
Number of samples . . . .	616	687	646	602	662	472
Median score for grade . .	9.2	10.7	11.0	11.3	12.2	12.8

Pupils writing above quality 12 should be excused from further drill, except voluntary drill.

**The Thorndike Scale.**—While it is assumed that the teacher will doubtless use the Ayres scale, because of its convenience and availability, yet teachers should know of the Thorndike Scale, and should appreciate the fact that it was Dr. E. L. Thorndike who first gave us a usable scale for handwriting.

The Thorndike scale is based upon general merit, as determined by the judgment of a large number of competent judges. In this respect it differs from the Ayres scale, which is based entirely upon legibility. It is unnecessary at this point to go into the discussion of the merits of the two scales.

<sup>1</sup> *Salt Lake City Survey*, page 149.

It is agreed that either scale can be understood, and will give much better results than the old method of grading. Because the Thorndike scale was first developed, and its value was immediately appreciated by school men, it was introduced into a large number of school systems, and is still retained in many of them. For this reason it will be well to indicate standards of quality according to the Thorndike scale. The numbers are quite definite since the samples on the Thorndike scale range from 4 to 18. The Thorndike scale was used in the Butte survey, and Table 13 shows the complete distribution of scores in quality.

TABLE 13. — THE DISTRIBUTION OF SCORES IN PENMANSHIP (Butte Survey, p. 165)

SCORE (QUALITY)	GRADE						
	2	3	4	5	6	7	8
0 . . . . .							
1 . . . . .							
2 . . . . .							
3 . . . . .							
4 . . . . .	5	2					
5 . . . . .	22	2	3	3		1	
6 . . . . .	21	21	16	3	2		1
7 . . . . .	29	44	24	12	1	3	3
8 . . . . .	28	86	42	56	20	15	7
9 . . . . .	42	41	55	61	25	29	15
10 . . . . .	7	8	20	16	9	11	1
11 . . . . .	29	13	21	17	32	25	23
12 . . . . .	5	2	15	15	44	12	21
13 . . . . .	7	2	2	6	17	19	9
14 . . . . .			3	4	10	16	9
15 . . . . .			1		9	6	15
16 . . . . .	1			1	10	12	17
17 . . . . .					6	2	3
18 . . . . .					3	1	
Total papers . . .	196	221	202	194	188	152	124
Median scores . . .	8.2	8.0	8.8	8.9	11.6	11.2	12.1

Table 14 shows the median performance in certain cities and indicates also the Freeman standard, expressed in units of the Thorndike scale.

TABLE 14. — QUALITY OF HANDWRITING (Thorndike)

	I	II	III	IV	V	VI	VII	VIII
Connersville, Indiana			10.3	10.0	10.3	11.7	11.7	11.0
Butte, Montana . .		8.2	8.0	8.8	8.9	11.6	11.2	12.1
Salt Lake City . .			9.2	10.7	11.1	11.3	12.2	12.8
Kansas City . . .	7.2	7.4	8.4	9.3	10.4	11.0	11.4	
Freeman's 56 cities <sup>1</sup>		7.8	9.4	10.2	11.2	11.6	12.5	13.4
Freeman's standard <sup>1</sup>		8.4	9.8	10.9	12.0	12.8	13.9	15.2

Table 15, which follows, contains a complete table of transformation, by which the qualities in the Ayres scale may be transformed into the Thorndike scale and vice versa.

TABLE 15. — COMPARATIVE VALUES.<sup>2</sup>

AYRES	THORNDIKE	THORNDIKE	AYRES
20	6.33	5	9.5
30	7.60	6	17.4
40	8.86	7	25.3
50	10.13	8	32.2
60	11.39	9	41.1
70	12.66	10	49.
80	13.93	11	56.9
90	15.19	12	64.8
		13	72.7
		14	80.6
		15	88.5
		16	96.4

<sup>1</sup> Transformed scores, approximate only.

<sup>2</sup> Dr. T. L. Kelley, *Journal of Educational Psychology*, December, 1914.

**Lister-Meyers Handwriting Scales.**—These scales are in use in the schools of Greater New York. They were prepared by Professors Lister and Meyers of the Brooklyn Training School for Teachers. They are printed on a sheet 24"×26" and show rankings from 90 to 20 on the three items: form, movement, and spacing. This scale is a good illustration of a special adaptation based upon the type of writing which the supervisors are endeavoring to secure in the particular city.

The teacher who is merely interested in putting her grading system on a scientific basis may neglect some of the present discussion and may secure good results by simply following the rules laid down for giving the tests, scoring the results, distributing the scores, and applying remedial instruction. What now seems theoretical and abstract in the measurement of handwriting will take on new significance as the teacher gradually masters the details of applying the work to her own schoolrooms. The practice will illuminate the theory; that which is theoretical will become practical. The work is of value as it modifies and improves school practice. Many teachers, however, will desire to know the history and development of the work, and in addition to a thorough study of the present chapter, will use the following bibliography to further study the subject.

#### BIBLIOGRAPHY

1. Ayres, L. P., "A Scale for Measuring the Quality of Handwriting of School Children," Russell Sage Foundation, Bulletin No. 113.
2. Freeman, F. N., "The Teaching of Handwriting," Riverside educational monographs, Houghton Mifflin Company.
- "Handwriting," the Fourteenth Yearbook of the National Society for the Study of Education, Chap. V, pp. 61-77.
- The Sixteenth Yearbook of the National Society for the Study of Education, Chap. IV, pp. 60-72.
- "An Analytical Scale for Judging Handwriting," *The Elementary School Journal*, April, 1915. Order copies from Houghton Mifflin Company, 25¢ each.

3. Thorndike, E. L., "Handwriting," *Teachers College Record*, March, 1910.  
"Teachers' Estimates of the Quality of Specimens of Handwriting," *Teachers College Record*, November, 1914.
4. Bobbitt, Franklin, Twelfth Yearbook of the National Society for the Study of Education, Part I, pp. 40-42.
5. Wilson, G. M., "The Handwriting by School Children," *Elementary School Teacher*, 1911, Vol. 11, pp. 540-543.
6. Kelley, T. L., *Journal of Educational Psychology*, December, 1914.
7. Gray, C. Truman, "Standard Score Card for Judging Handwriting," University of Texas, Austin, Texas.
8. Starch, Daniel, and Wise, Carl T., "A Measuring Scale for Handwriting." For copies, address The University Coöperative Co., Madison, Wisconsin. For discussion of the experimental and statistical work involved, see Starch, Daniel, "A Scale for Measuring Handwriting," *School and Society*, January, 1919.



## CHAPTER IV

### THE MEASUREMENT OF ARITHMETIC

MEASUREMENT in arithmetic is not so simple as in spelling or in handwriting. Arithmetic taken as a whole involves many processes, and each process in turn involves particular difficulties. No one has attempted to measure general mathematical ability,<sup>1</sup> and no one has attempted a test which covers the entire range of arithmetical processes. However, in its more essential and mechanical phases, arithmetic is susceptible of quite definite measurement. Since our purpose is to help the teacher in measuring her work to the extent that scales and standardized tests are available, it will be necessary to consider what can be measured at the present time, and what are the means available for doing it.

There are at the present time five series of tests reasonably well standardized. The one first developed and most extensively in use is the Courtis Standard Research Tests. At the present time Courtis is confining his work to measurement of the four fundamental processes by his tests known as "Series B."

The Courtis tests were first made available during the school year of 1909-10. The first tests, known as "Series A," were tentative in nature, and have since been discontinued. The use of Series B has extended very rapidly. During the

<sup>1</sup> The discovery of mathematical ability among secondary pupils has recently been attempted and a prognostic test devised. See Rogers, Agnes Low: "Experimental Tests of Mathematical Ability and Their Prognostic Value."

year 1915-16 they were used in forty-two states of the Union, Hawaii, and two foreign countries, a total of one half million copies being sold. The teacher will see, therefore, that in making use of the Courtis tests she is becoming acquainted with a method of measurement which is used widely and is likely to be used even more extensively in the future, unless they are replaced by something better. The chief attention of the teacher will be directed to these tests.

**Courtis Arithmetic Tests, Series B.**—Series B consists of tests in addition, subtraction, multiplication, and division, respectively. The test in addition consists of twenty-four examples, each made up of nine three-place numbers. They are constructed mechanically in such a way that each example is equal in difficulty to every other. The addition problems, therefore, consist of 24 different units of measurement, more or less in the nature of 24 foot rules, although not as accurately equal one to the other. The point of the test is to see how many of these examples can be solved by a pupil in a given time,—the time in this case being eight minutes. The examples in subtraction, multiplication, and division are likewise made up on the basis of uniform difficulty and a definite time limit is set. The details, which follow herewith, give a few samples from each of the tests, together with the time limits.



ARITHMETIC. TEST NO. 1. ADDITION

SCORE

Series B Form 2

No. Attempted.....

No. Right.....

You will be given eight minutes to find the answers to as many of these addition examples as possible. Write the answers on this paper directly underneath the examples. You are not expected to be able to do them all. You will be marked for both speed and accuracy, but it is more important to have your answers right than to try a great many examples.

127	996	237	386	186	474	877	537
375	320	949	463	775	787	845	685
953	778	486	827	684	591	981	452
333	886	987	240	260	106	693	904
325	913	354	616	372	869	184	511
911	164	600	261	846	451	772	988
554	897	744	755	595	336	749	559
167	972	195	833	254	820	256	127
<u>554</u>	<u>119</u>	<u>234</u>	<u>959</u>	<u>137</u>	<u>533</u>	<u>258</u>	<u>323</u>

## TEST NO. 2. SUBTRACTION

The test consists of 24 examples like the following, time 4 minutes.

97089301	93994413	108051861	163130569
<u>20203267</u>	<u>54783938</u>	<u>73463849</u>	<u>91061255</u>
168354186	188545364	120981427	105755782
<u>70537861</u>	<u>92471259</u>	<u>64188045</u>	<u>90863147</u>

## TEST NO. 3. MULTIPLICATION

The test consists of 25 examples like the following, time 6 minutes.

6283	9624	7853	4926	5873
<u>47</u>	<u>503</u>	<u>35</u>	<u>620</u>	<u>49</u>
2964	8357	6249	3785	4965
<u>94</u>	<u>87</u>	<u>78</u>	<u>35</u>	<u>19</u>

## TEST NO. 4. DIVISION

The test consists of 24 examples like the following, time 8 minutes.

29)24679	57)51642	38)32300	64)61504
46)34086	75)55500	92)27784	83)26643

**Nature of the Examples.** — The teacher may properly question whether the examples appearing in these tests are such as commonly appear in ordinary business transactions. That the examples of the Courtis test are more or less artificial from a social standpoint, and considerably more difficult than the transactions actually occurring under business conditions, is borne out by a study of the social and business use of arithmetic reported in Chapter 8 of the Sixteenth Yearbook of the National Society for the Study of Education.<sup>1</sup> The teacher should remember, however, that the purpose of the Courtis examples is merely to test the ability of the pupils in the fundamentals. Examples as difficult as those appearing in Test 1, for instance, will involve all of the difficulties of simpler examples, and so in a measure justify themselves in that they test the extreme ability likely to be required not only of pupils in the public schools under any system, but even by the exigencies of business and social situations of adult life. Mr. Courtis himself has recognized the fact that the multiplication and division problems are entirely too difficult for third grade pupils, and as a result his 1916 standards indicate a zero score for third grade pupils in multiplication and division. It is possible that in time further adjustment will be made in this same direction.

**Directions for Giving the Tests.** — It is assumed that the teacher is not interested in a merely theoretical discussion of arithmetic tests, but in using them for the measurement of the work in her own schoolroom. The next step, therefore, in connection with the Courtis tests is to write<sup>2</sup> for a sufficient quantity of the research tests in arithmetic, series B, to enable her to give the test to the number of pupils which she has in her room.

<sup>1</sup> This study has since been confirmed by a larger study. Wilson, G. M.: "A Survey of the Social and Business Usage of Arithmetic," Teachers College Bureau of Publications.

<sup>2</sup> See bibliography for directions.

The teacher will have no difficulty in administering the tests. One or two of the tests can be given during a single recitation period. The time for addition is 8 minutes, for subtraction 4 minutes. It will doubtless be better to give these two tests on one day, deferring the tests in multiplication and division until the next day. The instructions follow herewith.

#### INSTRUCTIONS TO EXAMINERS

1. For each room, prepare as many bundles of papers as there are rows of seats, putting into each bundle as many papers as there are seats in each row.

2. Begin by saying, "My purpose this morning is to measure how well this school teaches its children how to add, subtract, multiply, and divide. I have here some printed tests. They are not examinations, because exactly these same tests are given to all the grades from the third through high school. They are also being given in other schools in this city, and in other cities all over the country. It is the school that is being examined to-day. If you treat the tests as though they were a game, you will enjoy them and do your best for the honor of your school. I am going to give each of you a set of these papers, but do not look at them until I tell you to do so. Will the boys and girls in the front seats please distribute them for me?"

3. Distribute the papers by putting a bundle on the first desk in each row and letting the children do the rest.

4. Have the children fill out the blanks at the top of the first page. Write the date in figures, and the time to the nearest half hour; thus: 9-25-1913-10:30.

5. Have the children read instructions for Test 1 aloud in concert.

6. "Now please listen closely. In these tests it is important that we all start at the same time and stop at the same time. We can do this easily, if you follow my instructions exactly. Lay your papers on your desks in position to work the examples, but close the cover with your left hand, keeping it between your thumb and finger, like this (illustrate), so that you can open it quickly



when I tell you to start. Take your pencil in your right hand, and when I say 'Get ready,' raise your pencil hand in the air as if you were going to ask a question. (Illustrate, by suiting the action to the words.) Then when I say 'Start,' you can bring your pencil down as you turn the cover back, and every one will start at the same time. When I say 'Stop,' I want you all to stop at once, and to raise your hands again so that I can see that you have stopped. Now I think we are ready to try the test."

When the second hand of the watch reaches the 55-second mark say "Get ready for the addition test. Hands up." Exactly at the 60 mark say "Start."

#### ALLOW EXACTLY EIGHT MINUTES

"Stop. Hands up." Make sure all have stopped. "Count how many examples you have finished, and write the number in the score card in the corner under the number attempted. Do not count examples you have begun but have not finished. Your score is the number of the examples you have finished. I am coming to your desk to see that you have written it in the right place."

7. Read the answers from an Answer Card (be sure the form number corresponds with that of the tests), and have the children check answers right or wrong, counting the number right, and writing it in their score cards.

8. In similar fashion give and score the other tests.

For Test 2, Subtraction, allow exactly FOUR minutes.

For Test 3, Multiplication, allow exactly SIX minutes.

For Test 4, Division, allow exactly EIGHT minutes.

9. Give Tests 1 and 2 the first day, and Tests 3 and 4 the next. All may be given at one time if desired.

**Scoring the Results.**—The teacher may save herself much work, and on the whole secure equally satisfactory returns if she has each child score his own paper at the close of the tests (or papers may be exchanged for scoring). This method has the added advantage of enlisting the interest of the children.



At the conclusion of the test the pupil should make his own graph, showing attempts and rights in the four fundamental processes, as per the example which follows.



**ARITHMETIC**  
**Series B**

Name John Smith, Boy boy or girl, Age last birthday 14  
 School Eighth St. Grade 8 Room 8  
 City \_\_\_\_\_ State \_\_\_\_\_ Date May 15

[illegible]

GRAPH

ADDITION		SUBTRACTION		MULTIPLICATION		DIVISION	
Attempts	Rights	Attempts	Rights	Attempts	Rights	Attempts	Rights
24	24	24	24	24	24	24	24
23	23	23	23	23	23	23	23
22	22	22	22	22	22	22	22
21	21	21	21	21	21	21	21
20	20	20	20	20	20	20	20
19	19	19	19	19	19	19	19
18	18	18	18	18	18	18	18
17	17	17	17	17	17	17	17
16	16	16	16	16	16	16	16
15	15	15	15	15	15	15	15
14	14	14	14	14	14	14	14
13	13	13	13	13	13	13	13
12	12	12	12	12	12	12	12
11	11	11	11	11	11	11	11
10	10	10	10	10	10	10	10
9	9	9	9	9	9	9	9
8	8	8	8	8	8	8	8
7	7	7	7	7	7	7	7
6	6	6	6	6	6	6	6
5	5	5	5	5	5	5	5
4	4	4	4	4	4	4	4
3	3	3	3	3	3	3	3
2	2	2	2	2	2	2	2
1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	0

*Instructions.* In each column mark the number that corresponds to your score for that column. Then with a ruler draw a line from each number so marked to the next. Draw a curve for the class scores in the same way, using a dotted line. By comparing the two curves you can tell how much your scores are above or below the class results.

This individual score sheet will appeal to children, and will be exceedingly serviceable in securing the necessary further

progress of the children. The teacher may find it worth while to arrange all of the score sheets in order of excellence by pinning them on a piece of burlap on the side of the room. The pupil's score (dotted line) appearing on the individual score sheet above is the record of an eighth grade pupil who is about average in ability.

In order that the teacher may get an intelligent view of the performance of her class, it will be necessary to make a distribution of the class scores, and it is suggested that this be made in such a way as to show both speed and accuracy. Table 16, which follows herewith, shows such a distribution for 35 eighth grade pupils.

TABLE 16. — SHOWING DISTRIBUTION OF 35 EIGHTH GRADE PUPILS IN SPEED AND ACCURACY IN SEPTEMBER. ADDITION, SERIES B  
Score in Examples Attempted. (Speed)

Score in Examples right. (Accuracy)	5	6	7	8	9	10	11	12	13	14	15	16	17	18	TOTALS	Median Rights. (Accuracy) 7 or 63.6%
0 . .																
1 . .																
2 . .																
3 . .																
4 . .			I	I											2	
5 . .				I	I										2	
6 . .					I	I	2								4	
7 . .					2	I	4	2	I						10	
8 . .					I		2	2		I					6	
9 . .						I	I	I	2						5	
10 . .								I		I					2	
11 . .									I		I				2	
12 . .											I	I			2	
13 . .												I				
14 . .													I			
15 . .																
Totals			I	2	5	3	9	6	4	2	2	I			35	

Median Attempts  
(Speed) — 11

The table is to be read as follows: Out of 35 eighth grade pupils taking the addition test, one had a score of 7 attempts and 4 rights, one had a score of 8 attempts and 4 rights, etc.

It will be observed by reading the totals at the bottom of the sheet that one pupil attempted seven examples, two attempted eight, etc. One pupil in the class attempted as many as sixteen examples. The median score for attempts is 11.

The next question is, what proportion of the examples attempted were solved correctly. Totals for this may be read from the right-hand column. It will be observed that the one pupil who attempted seven problems solved four of them correctly. Of the two pupils who attempted eight, one solved four correctly, the other five. The one pupil who attempted 16 problems solved twelve of them correctly. The median performance of the group was seven problems solved correctly, and this is 63.6% of the median for attempts. It is quite evident from a study of the distribution that there is a wide range of ability in the class. The two pupils who solved twelve problems correctly certainly did three times as well as the two pupils who solved four problems correctly. Naturally the next question is, what is a reasonable standard for this particular eighth grade, and likewise for each of the grades.

**Standard Scores.**—The question is best answered by indicating the standards set up by Courtis, and by noting the performance of children throughout the country. The standards set up by Courtis are based upon an experience of five years and the scoring of thousands of test papers. In general, Courtis has placed his standards slightly above the median performance of children throughout the United States. The standards are indicated in Table 17, which follows herewith, and the median performances are shown in Table 18.

TABLE 17. — COURTIS STANDARD SCORES, FOR MAY, ARITHMETIC TESTS, SERIES B

GRADE	ADDITION		SUBTRACTION		MULTIPLICATION		DIVISION	
	1915	1916	1915	1916	1915	1916	1915	1916
III . . . . .	3	4	4	5	3	0	2	0
IV . . . . .	5	6	6	7	5	6	4	4
V . . . . .	7	8	8	9	7	8	6	6
VI . . . . .	9	10	10	11	9	9	8	8
VII . . . . .	11	11	11	12	10	10	10	10
VIII . . . . .	12	12	12	13	11	11	11	11

Standard of Accuracy 100%

The standards are given in Table 17, for the year 1915 as well as for 1916. The fact that after four or five years of work Courtis is still changing his standards tends to give the teacher a feeling of security in case her own results vary considerably from the standards. In fact, the individual teacher should not worry greatly about the standards, although they will be found valuable, and an attempt should be made to reach them. The work most helpful to the teacher, however, is the distribution of the scores for her own children, and the next section will discuss the significance of such distribution for a class.

The teacher who is interested in knowing what the performance of children elsewhere has been, is directed to Table 18. This gives the median performance of thousands of children for grades 4 to 8 inclusive throughout the country. Under each grade the first line gives the median as determined by Courtis in 1916, and on which he has based his 1916 standards. The second line under each grade gives the same data for 1915. The next four lines under each grade give the 1915 performance as reported from Indiana, Iowa, Kansas, and Boston, respectively. It is suggested that the



teacher study this table with particular reference to the grade or grades included in her own room. Is a speed of 11 and an accuracy of 63.6%, as shown for the eighth grade in Table 16, up to the average of children throughout the country? How much better is it than the average eighth grade child did in Indiana in 1915? How much below the Boston median performance? The answer to these questions will interest the children fully as much as they interest the teacher.

With the above questions in mind, and with the standards as given in Table 17, and the median performances as given in Table 18 before her, the teacher may now return to the distribution of grades for the 35 eighth grade pupils shown in Table 16. The median attempts for the 35 pupils are 11. The Curtis standard in addition for 1915 and 1916 is 12. It appears, therefore, that the median performance of the class is one below standard, so far as speed is concerned. It will be observed, however, that Boston alone, as shown in Table 18, equals or exceeds the Curtis standard. Curtis's own summary for performance in 1916 shows 10.2 as the median for the eighth grade (Table 18). It appears, therefore, that this particular eighth grade has done a little better than the eighth grade pupils reported on by Curtis in 1915 or 1916, and also a little better than the eighth grade pupils included in the summaries for Indiana, Iowa, or Kansas. In view of this comparison, the teacher may feel that her grade is on a reasonably satisfactory basis so far as speed is concerned. Turning now to accuracy, it is observed that this particular eighth grade class has an accuracy of 63.6%. Curtis calls for 100%. This, however, was not attained by any of the five groups included in Table 18. In addition the Boston eighth grade did best with an accuracy of 78%. It will be observed, however, that no one of the five groups goes quite as low as this particular eighth grade. This fact and the scattered distribution noted above emphasize the importance of giving attention to the addition work of this particular eighth grade.



TABLE 18. — MEDIAN PERFORMANCES FOR MAY, COURTIS ARITHMETIC TESTS, SERIES B

GRADE		ADDITION		SUBTRACTION		MULTIPLICATION		DIVISION		
		Speed	Accu- racy %	Speed	Accu- racy %	Speed	Accu- racy %	Speed	Accu- racy %	
IV.	Courtis .	1916	6.7	60	7.3	76	6.3	68	4.5	59
	Courtis .	1915	5.9	40	6.2	52	5.0	50	3.6	42
	Indiana .	1915								
	Iowa .	1915	6.9	58	7.4	73	6.3	66	4.8	63
	Kansas .	1915	5.9	51	6.4	64	5.2	58	3.8	56
	Boston .	1915	8.0	67	7.6	82	6.2	67	4.8	65
V.	Courtis .	1916	7.8	68	8.6	82	7.5	76	5.7	77
	Courtis .	1915	6.3	58	7.8	73	6.8	68	5.4	65
	Indiana .	1915	7.2	59	7.5	71	6.0	61	5.0	65
	Iowa .	1915	8.2	64	9.0	78	7.6	74	6.2	81
	Kansas .	1915	7.0	61	7.9	75	7.0	69	4.9	68
	Boston .	1915	9.4	71	9.3	85	7.7	76	6.5	81
VI.	Courtis .	1916	8.9	71	9.7	84	8.8	79	7.8	87
	Courtis .	1915	8.4	67	9.2	82	7.9	76	7.1	83
	Indiana .	1915	8.3	64	8.7	77	7.5	68	6.1	79
	Iowa .	1915	8.8	75	9.9	89	8.8	75	7.6	84
	Kansas .	1915	8.1	65	9.1	81	8.1	77	6.5	84
	Boston .	1915	11.1	75	11.1	87	9.4	79	8.7	87
VII.	Courtis .	1916	9.8	72	11.2	85	10.0	80	9.6	91
	Courtis .	1915	9.2	67	10.6	83	9.0	78	8.1	83
	Indiana .	1915	8.9	64	9.9	80	8.5	71	7.8	84
	Iowa .	1915	9.4	70	11.0	83	10.4	79	9.0	89
	Kansas .	1915	8.7	67	10.0	83	9.0	78	9.3	87
	Boston .	1915	12.3	76	12.2	87	10.5	81	10.2	90
VIII.	Courtis .	1916	10.2	74	12.1	85	11.0	81	10.4	93
	Courtis .	1915	10.2	67	12.3	81	10.6	82	10.6	91
	Indiana .	1915	9.5	67	10.9	82	9.9	74	9.7	87
	Iowa .	1915	10.3	72	12.9	86	11.6	83	12.0	93
	Kansas .	1915	9.8	71	11.5	86	10.9	82	10.9	92
	Boston .	1915	13.7	78	13.6	89	11.6	83	12.2	92

**Remedial Instruction.** — In discussing remedial instruction it will be well to recur again to the distribution of the eighth grade pupils shown in Table 16. The fact that no pupil reaches 100% accuracy indicates that the addition combinations are not fully mastered by any member of the class, that the addition problems in the test are too difficult for the mental attainments of the group, or that pupils are careless in their procedure. There may be other explanations of inaccuracy. Let us assume, however, that the first reason is the one which obtains with this particular grade, and it is quite likely the correct reason for most members of the class, since we may assume that an eighth grade class should be able to solve problems as difficult as those used in the test, and that if the test is administered properly, carelessness will not be very evident in the returns. The question, therefore, is how to re-teach the number combinations in such a way that they will be thoroughly known by every member of the class. It is not the province of the present work to discuss methods in any extended way, but merely to show the use of standard tests. If this test has revealed the defect correctly, it is then the teacher's problem to become acquainted with the methods which will enable her to teach the addition combinations. In brief, one of the best courses in arithmetic<sup>1</sup> indicates three steps in teaching the addition combinations: first, the mastery of the 45 elementary combinations and their reverses; second, carrying these same combinations up through the decades and drilling on the same until proficiency is obtained; and third, column addition.

If the teacher does not understand the details involved in these three steps, she will of course need to become acquainted with them. In discovering exactly what a particular pupil's difficulty is, the teacher will find it exceedingly valuable to have the pupil take one of the test examples at the board and

<sup>1</sup> The Connersville Course of Study in Elementary Mathematics. Republished by Warwick and York (in press).

proceed with his work orally. This will enable the teacher to follow and accurately observe the pupil's mental processes.

The above discussion emphasizes the fact that drill alone is not the most important consideration. The first duty of the teacher is to discover the difficulties of individual pupils. Then pupils can be grouped according to common difficulties. In this connection it may be mentioned that the Boston score, which stands well at the top, is a result obtained after careful procedure in diagnosis and correction, followed by needed drill, according to directions similar to the above, for a period of three years. Equally satisfactory results have been obtained in other cities where superior skill has directed the work in the mechanical phases of the fundamental processes. For example, the results obtained in an Indiana city<sup>1</sup> under a teacher who helped in making the Connersville course of study in arithmetic and was interested in the first use of the Curtis tests in that city, are not only much above the Indiana median, but they are even above the Boston average. These results were obtained by (1) systematizing the drill for the class as a whole, and (2) discovering the difficulties of individual pupils and giving the necessary specific help. All agree that drill, to be effective, must be intelligently systematized, and given at frequent intervals.

**Retesting.**—After the teacher has worked with her pupils faithfully, as individuals and as a class, she will want to retest the class in order to measure the results of her efforts. This may be done at any time, and the results will interest the members of the class fully as much as the teacher. The rules must be observed carefully in order that the test may be real and in order that comparisons may be valid.

Table 19 shows the results of a retest of the 35 pupils

Table 16, after four months of careful work. The  
that the pupils are still widely scattered in  
at the class as a whole has made good improve-

<sup>1</sup> Connersville, Indiana.

ment. The median of attempts has been raised from 11 to 12, and of rights from 7 to 10. The improvement in accuracy from 63.6% to 83.3% is very satisfactory. Because of individual differences, a teacher may expect wide variations in speed within an eighth grade class, but she should not be

TABLE 19. — SHOWING DISTRIBUTION OF THE 35 EIGHTH GRADE PUPILS OF TABLE 16, IN SPEED AND ACCURACY IN ADDITION, AFTER SPECIAL HELP AND DRILL (January)

Score in Examples Attempted (Speed)

	5	6	7	8	9	10	11	12	13	14	15	16	17	18	TOTALS	Median Rights. (Accuracy) 10 or 83.3%
0 . .																
1 . .																
2 . .																
3 . .																
4 . .																
5 . .																
6 . .																
7 . .				1	1										2	
8 . .					1	1	1								3	
9 . .						1	2	1	1						5	
10 . .						1	2	2	2	1					8	
11 . .							1	3	2						6	
12 . .								1	2	2	1				6	
13 . .									1	1	1	1			4	
14 . .												1			1	
15 . .																
Totals				1	2	3	6	7	8	4	2	2				

Median Attempts  
(Speed) — 12

content until pupils in an upper grade are letter perfect in solving simple examples in the fundamental processes, *i.e.* until 100% accuracy is reached.

Since the results in addition as shown in Table 16 were secured in September, and the results in Table 19 in January,

and since the standards as brought out in Tables 17 and 18 are based upon May tests, the teacher of the grade whose results are shown in Tables 16 and 19 may reasonably expect that her pupils will be up to standard when she gives the tests in May.

In Figure 6 the class improvement in rights from September to January is shown graphically. This graph comes directly from the rights in Tables 16 and 19.

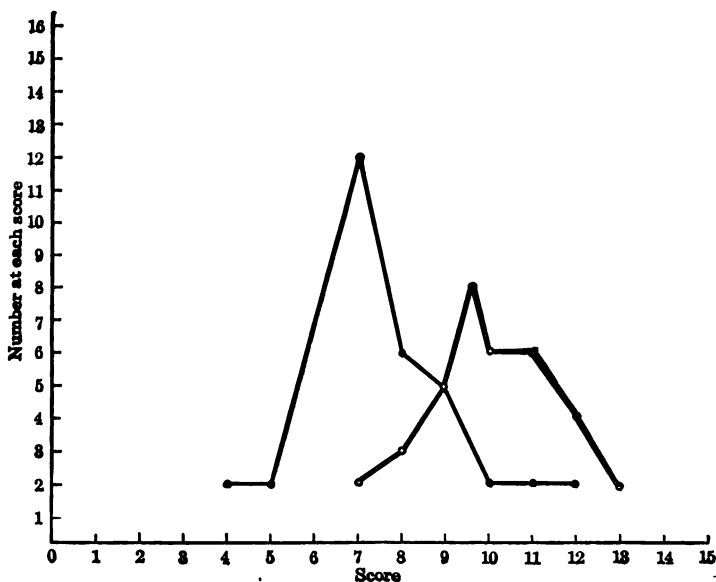


FIG. 6. — Showing attainment of the class in September (single line graph) and in January (double line graph). The entire class has moved steadily to the right, which means an improvement in score.

**Arithmetic Tests.** — Other available tests at the  
 are the Stone Reasoning Tests, the Woody  
 , the Boston Tests in Addition of Fractions,  
 Survey Arithmetic Tests, and the Monroe  
 in Arithmetic. These tests will in turn be



described briefly, more particular attention being given to the Stone Reasoning Tests and the Woody Arithmetic Scales, as these are needed to supplement the Curtis tests, and can be used by teachers without particular difficulty. It is evident that what can be measured in arithmetic depends somewhat on the test being used. In general it is the performance of the pupils which is tested, and this may include speed and accuracy, or accuracy only, according to how the tests are administered. The purpose in arithmetic, as in all testing, should be to find out the present condition of the child, in order to prescribe remedies in case he needs help, or in order to release him from further drill, in case he is fully up to reasonable standards.

**Reasoning Tests.**—When arithmetic is put to practical business use, it is always connected with an actual situation, and the solution requires judgment or reasoning as to the processes involved. For upper grade work no test of arithmetic is complete which fails to test reasoning ability. The Stone Reasoning Test has been most used. It consists of twelve problems, ranging in value from 1 to 2, as follows:

### THE STONE REASONING TEST

(Time Exactly 15 minutes)

School.....Grade.....Name of pupil.....

PROBLEM VALUE	PROBLEMS
	Solve as many of the following problems as you have time for; work them in order as numbered:
1.0	1. If you buy 2 tablets at 7 cents each and a book for 65 cents, how much change should you receive from a two-dollar-bill?
1.0	2. John sold 4 <i>Saturday Evening Posts</i> at 5 cents each. He kept $\frac{1}{2}$ the money and with the other $\frac{1}{2}$ bought Sunday papers at 2 cents each. How many did he buy?



THE STONE REASONING TEST (*Continued*)

PROBLEM VALUE	PROBLEMS
1.0	3. If James had 4 times as much money as George, he would have \$16. How much money has George?
1.0	4. How many pencils can you buy for 50 cents at the rate of 2 for 5 cents?
1.0	5. The uniforms for a baseball nine cost \$2.50 each. The shoes cost \$2 per pair. What was the total cost of uniforms and shoes for the nine?
1.4	6. In the schools of a certain city there are 2200 pupils; $\frac{1}{3}$ are in the primary grades, $\frac{1}{4}$ in the grammar grades, $\frac{1}{5}$ in the high school and the rest in the night school. How many pupils are there in the night school?
1.2	7. If $3\frac{1}{2}$ tons of coal cost \$21, what will $5\frac{1}{2}$ tons cost?
1.6	8. A news dealer bought some magazines for \$1. He sold them for \$1.20, gaining 5 cents on each magazine. How many magazines were there?
2.0	9. A girl spent $\frac{1}{3}$ of her money for car fare and three times as much for clothes. Half of what she had left was 80 cents. How much money did she have at first?
2.0	10. Two girls receive \$2.10 for making buttonholes. One makes 42, the other 28. How shall they divide the money?
2.0	11. Mr. Brown paid one third of the cost of a building; Mr. Johnson received \$500 more annual rent than Mr. Brown. How much did each receive?
2.0	12. A freight train left Albany for New York at 6 o'clock. An express train left on the same track at 8 o'clock. It went at the rate of 40 miles an hour. At what time of day will it overtake the freight train if the freight train stops after it has gone 56 miles?

The papers are scored by giving to each problem solved correctly the value as indicated at the left of each problem in the above. The test was first formulated for upper sixth grade pupils, but it is equally good for seventh or eighth

grade pupils. It is too difficult for good results in grades below the sixth.

Dr. Stone has recently issued<sup>1</sup> the following grade standards:

GRADE	STANDARD
5	Score of 5.5, reached or exceeded by 80%, 75% accuracy.
6	Score of 6.5, reached or exceeded by 80%, 80% accuracy.
7	Score of 7.5, reached or exceeded by 80%, 85% accuracy.
8	Score of 8.75 reached or exceeded by 80%, 90% accuracy.

It is quite probable that the median scores secured through the use of the Stone reasoning tests in various surveys form a more usable standard than the one suggested by Dr. Stone. These scores are shown in Table 20.

TABLE 20. — SHOWING MEDIAN SCORES OBTAINED IN THE USE OF THE STONE REASONING TESTS

GRADE	STONE 1908 26 CITIES	BUTTE, MONT. 1914	SALT LAKE CITY 1915	BOSTON 1916	BROOKLINE MASS.	LEAD S. D.	NASSAU CO. N. Y. 1917 <sup>2</sup>
5		2.2	3.7		4.0		
6	5.5	3.9	6.4	4.0	6.2	6.7	4.5
7		5.8	8.6	6.4			
8		7.7	10.5			11.6	7.2

The teacher will find it worth while to use the Stone reasoning tests, although the standards are not so definite as for the Courtis tests in the fundamentals. It will be simpler to take the returns from a single city, as for example, Salt

<sup>1</sup> Stone, C. W., "Standardized Reasoning Tests in Arithmetic and How to Use Them." (Teachers College Bureau of Publications.)

<sup>2</sup> The scoring is such as to slightly raise the score.

Lake City, as a standard. If pupils fail to reach the Salt Lake City standard, they are not doing as well as pupils have done in an average city system.

**Diagnostic Tests.** — The teacher who has followed the discussion closely will appreciate the fact that the Courtis tests, while measuring ability, do not analyze the difficulties and do not permit the teacher to use them easily for analyzing a pupil's shortcomings. This defect of the Courtis tests is being overcome gradually by the formation of other tests which have better diagnostic possibilities. Among these are the Woody Arithmetic Scales.

**Woody Scales.** — The Woody scales were not originally designed for diagnostic purposes, but they are being made to serve that purpose, as well as their original purpose of measuring the ability of children. They are constituted quite differently from the Courtis tests. Each Courtis test consists of a series of problems of equal difficulty in one of the fundamental processes. The Woody scales consist of a series of problems of increasing difficulty. They are designed to measure work in the four fundamental operations: addition, subtraction, multiplication, and division. While constructed on a statistical basis rather than for the purpose of serving as the basis of an analysis of subject matter needs, yet at the same time they do cover subject matter reasonably well. The addition scale, for instance, covers simple combinations in one, two, three, and four column addition; examples with addends from two to sixteen; addition of simple fractions; addition of decimals; addition of U. S. money; addition of denominate numbers; and addition of mixed numbers. The additions are expressed in column form and by the plus sign. Thus the pupil is tested, more or less, over the entire range of addition possibilities, by a series of problems ranging in difficulty from those so simple that any third grade pupil may solve them, up to other problems so difficult that few eighth grade pupils succeed in solving them. This will appear by exami-

nation of the addition scale which follows herewith. The subtraction, multiplication, and division scales also follow.

SERIES A <sup>1</sup>

ADDITION SCALE

(20 minutes)

Name.....  
When is your next birthday?.....How old will you be?....  
Are you a boy or girl?.....In what grade are you?.....

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2	2	17	53	72	60	3+1=	2+5+1=	20
<u>3</u>	4	<u>2</u>	<u>45</u>	<u>26</u>	<u>37</u>			10
	<u>3</u>							2
								30
								<u>25</u>

(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
21	32	43	23	24+42=	100	9	199	2563
33	59	1	25		33	24	194	1387
<u>35</u>	<u>17</u>	2	<u>16</u>		45	12	295	4954
		<u>13</u>			201	15	<u>156</u>	<u>2065</u>
					<u>46</u>	<u>19</u>		

(19)	(20)	(21)	(22)	(23)	(24)	(25)
\$ .75	\$12.50	\$8.00	547	$\frac{1}{8} + \frac{1}{8} =$	4.0125	$\frac{3}{8} + \frac{5}{8} + \frac{7}{8} + \frac{1}{8} =$
1.25	16.75	5.75	197		1.5907	
<u>.49</u>	<u>15.75</u>	2.33	685		4.10	
		4.16	687		<u>8.673</u>	
		.94	456			
		<u>6.32</u>	393			
			525			
			240			
			<u>152</u>			

<sup>1</sup> The scales are printed in large type, on separate sheets, 8½"×11", with ample space for the insertion of answers.

(26)	(27)	(28)	(29)	(30)	(31)	(32)
$12\frac{1}{2}$	$\frac{1}{8} + \frac{1}{4} + \frac{1}{2} =$	$\frac{3}{4} + \frac{1}{4} =$	$4\frac{3}{4}$	$2\frac{1}{2}$	113.46	$\frac{3}{4} + \frac{1}{2} + \frac{1}{4} =$
$62\frac{1}{2}$			$2\frac{1}{4}$	$6\frac{3}{8}$	49.6097	
$12\frac{1}{2}$			<u><math>5\frac{1}{4}</math></u>	$3\frac{3}{4}$	19.9	
<u><math>37\frac{1}{2}</math></u>					9.87	
					.0086	
					18.253	
					<u>6.04</u>	

(33)	(34)	(35)	(36)	(37)
.49	$\frac{1}{8} + \frac{3}{8} =$	2 ft. 5 in.	2 yr. 5 mo.	$16\frac{1}{8}$
.28		3 ft. 5 in.	3 yr. 6 mo.	$12\frac{1}{8}$
.63		<u>4 ft. 9 in.</u>	4 yr. 9 mo.	$21\frac{1}{2}$
.95			5 yr. 2 mo.	<u><math>32\frac{3}{4}</math></u>
1.69			<u>6 yr. 7 mo.</u>	
.22				
.33				
.36				
1.01				
.56				
.88		(38)		
.75		$25.091 + 100.4 + 25 + 98.28 + 19.3614 =$		
.56				
1.10				
.18				
<u>.56</u>				

## SERIES A

## SUBTRACTION SCALE

Name.....

When is your next birthday?.....How old will you be?....

Are you a boy or girl?.....In what grade are you?.....

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
8	6	2	9	4	11	13	59	78	$7 - 4 =$	76
<u>5</u>	<u>0</u>	<u>1</u>	<u>3</u>	<u>4</u>	<u>7</u>	<u>8</u>	<u>12</u>	<u>37</u>	<u>      </u>	<u>60</u>

(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
27	16	50	21	270	393	1000	567482	$2\frac{3}{4} - 1 =$
<u>3</u>	<u>9</u>	<u>25</u>	<u>9</u>	<u>190</u>	<u>178</u>	<u>537</u>	<u>106493</u>	

(21)	(22)	(23)	(24)	(25)	(26)
10.00	$3\frac{1}{2} - \frac{1}{2} =$	80836465	$8\frac{7}{8}$	27	4 yd. 1 ft. 6 in.
<u>3.49</u>		<u>49178036</u>	<u><math>5\frac{3}{4}</math></u>	<u><math>12\frac{5}{8}</math></u>	<u>2 yd. 2 ft. 3 in.</u>

(27)	(28)	(29)	(30)
5 yd. 1 ft. 4 in.	$10 - 6.25 =$	$75\frac{3}{4}$	$9.8063 - 9.019 =$
<u>2 yd. 2 ft. 8 in.</u>		<u><math>52\frac{1}{4}</math></u>	

(31)	(32)	(33)	(34)	(35)
$7.3 - 3.00081 =$	1912 6 mo. 8 da.	$\frac{5}{12} - \frac{2}{10} =$	$6\frac{1}{8}$	$3\frac{7}{8} - 1\frac{5}{8} =$
	<u>1910 7 mo. 15 da.</u>		<u><math>2\frac{7}{8}</math></u>	

# SERIES A

## MULTIPLICATION SCALE

Name.....  
 When is your next birthday?.....How old will you be?....  
 Are you a boy or girl?.....In what grade are you?.....

(1)	(2)	(3)	(4)	(5)	(6)	(7)
$3 \times 7 =$	$5 \times 1 =$	$2 \times 3 =$	$4 \times 8 =$	23	310	$7 \times 9 =$
				<u>3</u>	<u>4</u>	

(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
50	254	623	1036	5096	8754	165	235
<u>3</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>6</u>	<u>8</u>	<u>40</u>	<u>23</u>

(16)	(17)	(18)	(19)	(20)	(21)	(22)
7898	145	24	9.6	287	24	$8 \times 5\frac{3}{4} =$
<u>9</u>	<u>206</u>	<u>234</u>	<u>4</u>	<u>.05</u>	<u><math>2\frac{1}{2}</math></u>	

(23)	(24)	(25)	(26)	(27)	(28)	(29)
$1\frac{1}{4} \times 8 =$	16	$\frac{7}{8} \times \frac{3}{4} =$	9742	6.25	.0123	$\frac{1}{8} \times 2 =$
	<u><math>2\frac{5}{8}</math></u>		<u>59</u>	<u>3.2</u>	<u>9.8</u>	



(30) $\begin{array}{r} 2.49 \\ \underline{36} \end{array}$	(31) $\frac{1\frac{1}{2}}{3} \times \frac{1\frac{5}{8}}{2} =$	(32) $\begin{array}{r} 6 \text{ dollars } 49 \text{ cents} \\ \underline{8} \end{array}$	(33) $2\frac{1}{2} \times 3\frac{1}{2} =$	(34) $\frac{1}{2} \times \frac{1}{2} =$
(35) $\begin{array}{r} 987\frac{3}{4} \\ \underline{25} \end{array}$	(36) $\begin{array}{r} 3 \text{ ft. } 5 \text{ in.} \\ \underline{5} \end{array}$	(37) $2\frac{1}{4} \times 4\frac{1}{2} \times 1\frac{1}{2} =$	(38) $\begin{array}{r} .0963\frac{1}{8} \\ \underline{.084} \end{array}$	(39) $\begin{array}{r} 8 \text{ ft. } 9\frac{1}{2} \text{ in.} \\ \underline{9} \end{array}$

## SERIES A

## DIVISION SCALE

Name.....

When is your next birthday?..... How old will you be?....

Are you a boy or girl?..... In what grade are you?.....

(1) $\begin{array}{r} 3 \overline{)6} \end{array}$	(2) $\begin{array}{r} 9 \overline{)27} \end{array}$	(3) $\begin{array}{r} 4 \overline{)28} \end{array}$	(4) $\begin{array}{r} 1 \overline{)5} \end{array}$	(5) $\begin{array}{r} 9 \overline{)36} \end{array}$	(6) $\begin{array}{r} 3 \overline{)39} \end{array}$
(7) $4 \div 2 =$	(8) $\begin{array}{r} 9 \overline{)0} \end{array}$	(9) $\begin{array}{r} 1 \overline{)1} \end{array}$	(10) $6 \times ? = 30$	(11) $\begin{array}{r} 2 \overline{)13} \end{array}$	(12) $2 \div 2 =$
(13) $\begin{array}{r} 4 \overline{)24 \text{ lb. } 8 \text{ oz.}} \end{array}$	(14) $\begin{array}{r} 8 \overline{)5856} \end{array}$	(15) $\frac{1}{4} \text{ of } 128 =$	(16) $\begin{array}{r} 68 \overline{)2108} \end{array}$	(17) $50 \div 7 =$	
(18) $\begin{array}{r} 13 \overline{)65065} \end{array}$	(19) $248 \div 7 =$	(20) $\begin{array}{r} 2.1 \overline{)25.2} \end{array}$	(21) $\begin{array}{r} 25 \overline{)9750} \end{array}$	(22) $\begin{array}{r} 2 \overline{)13.50} \end{array}$	
(23) $\begin{array}{r} 23 \overline{)469} \end{array}$	(24) $\begin{array}{r} 75 \overline{)2250300} \end{array}$	(25) $\begin{array}{r} 2400 \overline{)504000} \end{array}$	(26) $\begin{array}{r} 12 \overline{)2.76} \end{array}$		
(27) $\frac{7}{8} \text{ of } 624 =$	(28) $\begin{array}{r} .003 \overline{).0936} \end{array}$	(29) $3\frac{1}{2} \div 9 =$	(30) $\frac{3}{4} \div 5 =$		
(31) $\frac{5}{4} \div \frac{3}{8} =$	(32) $9\frac{5}{8} \div 3\frac{3}{4} =$	(33) $\begin{array}{r} 52 \overline{)3756} \end{array}$			
(34) $62.50 \div 1\frac{1}{4} =$	(35) $\begin{array}{r} 531 \overline{)37722} \end{array}$	(36) $\begin{array}{r} 9 \overline{)69 \text{ lb. } 9 \text{ oz.}} \end{array}$			

**Directions for Giving Woody Tests.**— The directions for administering the Woody scales accompany the sheets, which may be secured from Teachers College Bureau of Publications. It is quite essential that uniform methods be followed in order to make results comparable. The papers are distributed with face down. When pupils are ready with pencils in hand, they are told to turn over the paper and answer the questions at the top of the page. The specific directions for the addition test as given by Dr. Woody are as follows: "Every problem on the sheet which I have given you is an addition problem, an 'and problem.' Work as many of these problems as you can and be sure you get them right. Do all of your work on this piece of paper and don't ask anybody any questions. Begin."

For the series A scales, twenty minutes are allowed for each test. There are shortened scales, series B, which are given in ten minutes each, but since the purpose in using the Woody scales will doubtless be to benefit more or less by their diagnostic values, it is assumed that teachers will prefer to use the longer scales of series A. It may be noted at this point that the time for giving the tests has been varied.<sup>1</sup> While a shortened time gives slightly better distributions, particularly in the upper grades, yet the problems at the upper end of the scales are so difficult that few pupils will solve them even when given all of the time necessary. As a matter of fact, 20 minutes, the time allowed, is sufficient for most upper grade pupils to complete any one of the tests. The result is that in the upper grades, accuracy only is measured. But in using the Woody scales, it is likely that accuracy is the thing in which the teacher will be chiefly interested.

In using the other Woody scales, the directions are the same as for addition except the substitution of such expressions as "subtract or 'take away' problems"; "multiplication or

<sup>1</sup> The Nassau County Survey used 18 minutes instead of 20.

TABLE 21. — ANSWERS TO PROBLEMS IN WOODY SCALES

PROBLEM	ADDITION	SUBTRACTION	MULTIPLICATION	DIVISION
1	5	3	21	2
2	9	6	5	3
3	19	1	6	7
4	98	6	32	5
5	98	0	69	4
6	97	4	1,240	13
7	4	5	63	2
8	8	47	150	0
9	87	41	1,524	1
10	89	3	4,361	5
11	108	16	8,288	$6\frac{1}{2}$ not $6+1$
12	59	24	30,576	1
13	64	7	70,032	6 lb. 2 oz. not $6+2$
14	67	25	6,600	732
15	425	12	5,405	32
16	79	80	71,082	31
17	844	215	29,870	$7\frac{1}{2}$ not $7+1$
18	10,966	463	5,616	5,005
19	\$2.49	460,989	38.4	$35\frac{1}{2}$ not $35+3$
20	\$45.00	$1\frac{1}{2}$	14.35	12
21	\$27.50	6.51	60	300
22	3,873	3	46	6.75
23	$\frac{1}{2}$	31,658,429	10	$20\frac{2}{3}$ ; 20.3, not $20+9$
24	18 3762	$3\frac{1}{2}$	42	30,004
25	2, not $\frac{1}{2}$ nor $\frac{1}{4}$	$14\frac{1}{2}$	$\frac{1}{2}$	210
26	125, not $123\frac{1}{2}=2$	1 yd. 2 ft. 3 in. not 63 in.	574,778	.23
27	$\frac{1}{2}$	2 yd. 1 ft. 8 in. not 81 in.	20,000	546
28	1 not $\frac{1}{2}$ nor $\frac{1}{4}$	$3\frac{1}{2}$ or 3.75	.12,054	31.2
29	$12\frac{1}{2}$ not $11\frac{1}{2}=1\frac{1}{2}$	$23\frac{1}{2}$ not $23\frac{1}{2}=\frac{1}{2}$	$\frac{1}{2}$ not $\frac{1}{4}$	$\frac{1}{18}$
30	$12\frac{1}{2}$ not $11\frac{1}{2}=1\frac{1}{2}$	.7873	89.64	$\frac{1}{2}$ or .15
31	217.1413	4.29919	$\frac{9}{16}$	$2\frac{1}{2}$
32	$1\frac{1}{2}$ not $\frac{1}{2}$ nor $1\frac{1}{4}=\frac{1}{2}$	1 yr. 10 mo. 23 da.	\$51.92 or 51 dol. 92 cts.	$2\frac{1}{2}$
33	10.55	$\frac{1}{2}$	$8\frac{1}{2}$	$72\frac{1}{3}$ or 72.23
34	$\frac{1}{2}$	$3\frac{1}{2}$ not $3\frac{1}{2}=\frac{1}{2}$	$\frac{1}{4}$	50
35	10 ft. 8 in. or $10\frac{2}{3}$ ft.	$2\frac{1}{2}$ not $2\frac{1}{2}=\frac{1}{2}$	24,693 $\frac{1}{2}$	$71\frac{1}{7}$ or 71.04
36	22 yr. 5 mo. or $22\frac{1}{3}$ yr.		17 ft. 1 in.	7 lb. $11\frac{1}{2}$ oz.
37	$82\frac{1}{2}$		$15\frac{1}{4}$	
38	268.1324		.0080902 $\frac{1}{2}$ or .00809025	
39			79 ft. $1\frac{1}{2}$ in.	

TABLE 22

NUMBER OF PROBLEMS SOLVED	SCORE	
	As Recorded	Totals
1		
2		
3		
4		
5	/	1
6	/	1
7		
8	////	4
9	///	2
10	/	1
11	///	3
12	////	6
13	///	2
14	//////	8
15	///	3
16	///	2
17	/	1
18	/	1
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		

Class Median 12

'times' problems"; "division or 'into' problems," for "addition or 'and' problems." In case the teacher has used other expressions to indicate one of the processes, these may be substituted for the expressions "and" problem, etc. The purpose in using these extra expressions is to make clear to the child the process which is involved in the particular test.

The standard for marking the examples in the Woody scales is absolute accuracy, and the final answer should be in its lowest terms. The table on the opposite page gives the correct list of answers.

The method of tabulating the results of the Woody tests is very simple. Assuming that a test has been given, indicate on the upper corner of each page the number of problems solved correctly. Then, for convenience, arrange the papers in order according to the number of problems solved. With the papers thus arranged, it will be possible to draw off directly the results of the test as shown in Table 22. This table shows the distribution of a class of 35 with reference to the number of problems solved by the different members.

Table 22 is taken directly from the results of a division test given in an intermediate grade,<sup>1</sup> November 1, 1917. The distribution of pupils' scores resulting from giving the Woody Test in Division, Series A, is shown for the entire school system in Table 23.

This is the same form as shown in Table 22, except that it covers five grades, and the number of pupils in each grade is the complete number for the entire city. The superintendent of this school system has shown, exceptionally well, the diagnostic possibilities of the Woody scales. In the study referred to, he analyzes the division difficulties of pupils as shown by the errors they have made in attempting to solve the problems in the division scale. It will be well for the

<sup>1</sup> Anderson, C. J., "Use of Woody Scales for Diagnostic Purposes," *Elementary School Journal*, 16: June, 1918, pp. 770-781.

TABLE 23. — DISTRIBUTION OF PUPILS' SCORES

NUMBER OF PROBLEM	GRADE				
	IV	V	VI	VII	VIII
1		1			
2	1				
3					
4					
5	1				
6	3	1			
7	1	1			
8	11	2			
9	7				
10	5	9	2		
11	12	7	3		
12	11	6			
13	8	9	1	1	
14	13	3	2		1
15	5	2	2		1
16	3	7	4		1
17	1	10	4		
18	2	2	4		
19		12	2	2	
20		6	8	2	2
21		9	6	1	
22		8	9	6	
23		2	12	1	
24		1	11	3	1
25		1	9	2	1
26		1	5	8	1
27			3	3	2
28			3	9	3
29			1	5	3
30				5	4
31				9	3
32			1	15	6
33				5	3
34			1	2	5
35				3	2
36				1	1
Total	84	100	93	83	40
Median	12	17	22	29	30.5



teacher to summarize for her class the number of wrong solutions for each problem attempted. This can be shown for a single test by a table similar to Table 22, in which the problems are listed by number on the left, and the number of incorrect solutions shown on the right. In the final analysis, however, the teacher should study each paper to see what mistakes each particular pupil made. This should be done in each of the fundamental processes. If the Woody scales are used to supplement the Curtis tests for the purpose of finding out where the pupil made his mistakes, it will be found exceedingly valuable. The various types of errors made in division in the city referred to were summarized by the superintendent and his teachers as follows:

1. Ignorance of multiplication tables, 30 per cent. Illustration:  $\begin{array}{r} 8,107 \\ 8 \overline{)5,856} \end{array}$

2. Using dividend as a whole, 14 per cent. Illustration:  $\begin{array}{r} 3 \overline{)39} \\ 12-3 \end{array}$

3. Confusion of multiplication and division, 14 per cent. Illustration:  $\begin{array}{r} 93 \\ 3 \overline{)39} \end{array}$

4. Remainder, 10 per cent. Illustration:  $\begin{array}{r} 6\frac{2}{3} \\ 2 \overline{)13} \end{array}$

5. Confusion of signs, 7 per cent. Illustration:  $2 \div 2 = 4$ .

6. Form of example strange, 5 per cent. Illustration:  $\frac{1}{4}$  of 128.

7. Carrying (either forgetting to carry or ignorance of what should be carried), 5 per cent. Illustration:  $\begin{array}{r} 1.350 \\ 2 \overline{)620} \end{array}$

8. Value of 0, 5 per cent. Illustration:  $\begin{array}{r} 9 \overline{)0} \quad 1 \overline{)1} \\ 9 \quad 0 \end{array}$

9. Confusion of addition and multiplication, 5 per cent. Illustration:  $\begin{array}{r} 3 \overline{)6} \\ 3 \end{array}$

10. Confusion of dividend and divisor, 2 per cent. Illustration:  $8 \overline{)498}$ .

212

(This quotient is explained as follows: 4 into 8 = 2, 8 into 9 1 and 1 over, 8 into 18 = 2 - 2 over.)

11. Using some figure in dividend twice, 2 per cent. Illustration:  $8 \overline{)5,856}$ .

7,107

12. Transposing answer, 1 per cent. Illustration:  $\frac{1}{4}$  of 128 = 23.

The teacher with her particular grade should proceed in a similar manner, taking up each fundamental process and discovering the types of errors made. It will be well to note, not only the errors made, but after each, the names of the pupils making that particular mistake in order that she may give special attention to all of the pupils making a particular mistake. Suppose, for example, that one of the teachers in the above city, on a certain day, desires to work upon the fourth one of the listed errors, namely *inability to handle the remainder*. In an intermediate class of 35 pupils, she may have four who need help on this point. By referring to her paper she will be able to call the names of the four who need special instruction. So with each particular mistake, she will be able to call for the pupils who need help, permitting others in the class to spend their time in some other way.

The superintendent and his teachers in the city referred to, noted that long division was difficult for the pupils, and so made a special summary of the errors in long division, as follows:

1. The assumption that the first integer of the divisor may be used always as a trial divisor.
2. The trial-and-error method of finding quotient.
3. Ignorance of multiplication tables.
4. Carrying wrong number when multiplying.
5. Borrowing in subtraction.
6. Ignorance of value of cipher.
7. Forgetting to place integers in quotient.

This is a good illustration of the diagnostic use of tests as a basis for remedial instruction. Such use of tests makes them of direct service in the work of helping pupils, and this is a use that must in future receive more and more attention.

The Woody tests are being quite extensively used. During the year 1917-1918, 300,000 copies of the tests were used by school men in the United States. This extensive use is gradually developing standards of performance. Instead of giving standards alone, it will be more helpful to list the returns from a number of cities. Accordingly there are given in Table 24 the median scores secured in the use of the Woody scales in twenty Wisconsin cities, as well as *the Woody standard medians*.

TABLE 24.—MEDIAN SCORES BY CITIES

WISCONSIN CITIES	DATE TESTED	SECTION TESTED	ADDITION						SUBTRACTION					
			III	IV	V	VI	VII	VIII	III	IV	V	VI	VII	VIII
1	10/3/16	B	12.4	19.3	21.2	27.4	30.3	31.9	10.7	16.3	20.3	24.4	27.7	30.8
2	10/10/16	B	12.7	17.8	23.1	27.	30.3	32.	11.9	17.3	22.3	27.4	29.4	30.5
3	10/24/16	B	12.6	17.7	23.7	28.8	30.6	31.6	13.8	13.3	21.9	27.8	26.8	31.5
4	2/6/17	A&B	15.5	21.4	22.7	30.7	33.3	34.3	13.6	18.3	20.7	25.7	27.6	30.6
5	2/12/17	A&B	20.7			28.3	32.1	32.8		18.5	19.9	25.		27.6
6	11/27/16	B	15.3	20.7	22.1	28.	31.4	33.4	14.9	18.5	21.3	28.5	30.2	32.8
7	12/5/16	B	13.	20.1	20.3	24.2	27.4	26.8	11.1	18.1	19.9	20.7	27.4	26.7
8	1/9/17	B	19.2	20.7	26.9	30.9	31.9	32.9	15.7	20.	22.7	26.9	30.6	32.6
9	1/12/17	B	14.8	19.8	25.3	27.	28.2	32.7						
10	12/5/16	A&B		18.5	20.9	27.	30.		11.3	17.3				29.
11	3/27/17	A&B	16.6	19.5	23.3	29.3	32.4	33.2	13.4	18.4	20.5	25.2	28.7	29.2
12	2/26/17	A	14.4	17.5	20.6	23.8	31.2	33.1	12.6	17.9	20.1	24.2	27.5	
13	3/8/17	A	12.5	19.	20.5	26.8	30.8	32.5						
14														
15	5/8/17	A&B							8.7	9.8	17.7	18.4	24.7	28.2
16	4/10/17	A							17.8	22.	23.9	29.4	30.5	30.2
17	4/12/17	A							12.	19.2				
18	6/4/17	A&B							12.2	18.2	21.2	27.3		
19	6/6/17	A&B		21.8	26.4	28.6	32.	31.8						
20	5/ /17	A	19.	22.7	26.8	30.8	33.9	34.8	15.	20.3	24.	28.8	32.9	33.
Median . . . . .			15.5	20.2	22.7	28.4	31.9	33.1	13.3	18.1	20.8	25.6	28.4	30.3
Woody's Standard Median			14.5	18.3	23.1	29.8	32.4	34.	11.2	15.7	20.4	25.	28.5	31.7

TABLE 24 (Continued)

MULTIPLICATION										DIVISION					
1	10/3/16	B	2.7	13.7	19.8	26.9	30.2	32.8		3.1	9.9	19.2	25.4	27.3	29.9
2	10/10/16	B		12.4	21.4	29.5	29.7	32.9			11.9	19.9	26.	29.	30.
3	10/24/16	B	4.3	11.7	18.8	27.2	29.3	32.3		5.6	5.5	17.8	25.2	28.2	30.2
4	2/5/17	A&B	6.2	14.3	18.8	28.5	30.5	33.6		8.	13.1	19.4	25.9	28.5	30.7
5	2/12/17	A&B			18.9	27.1									
6	11/27/16	B	6.6	17.8	18.9	28.4	31.	34.			13.3	20.5	27.6	31.	29.9
7	12/5/16	B	8.0	12.8	16.1	18.6	27.	29.6			10.8	10.8	17.6	26.5	25.5
8	1/9/17	B	8.5	15.8	18.9	28.3	32.1	34.1			12.8	18.1	24.3	28.8	31.1
9	1/12/17	B					30.4				14.3	16.	25.4	24.6	27.7
10	12/5/16	A&B	5.1	18.5		26.3					10.4		18.5	22.1	22.5
11	3/27/17	A&B	7.8	16.4	19.2	27.3	32.3	32.4			11.6	16.5	23.2	27.5	28.2
12	2/26/17	A		15.5	18.4	24.5	30.4	31.7			11.8	16.9	20.2	29.	28.7
13	3/8/17	A									11.6	15.2	21.5	23.5	25.2
14	5/4/17	A&B		20.5	23.5	30.5	32.4	35.							
15	5/8/17	A&B								8.7	9.8	17.7	18.4	24.7	28.2
16	4/12/17	A			22.5	24.5								23.5	23.3
17	6/4/17	A&B								14.8			25.	29.5	29.4
18	6/6/17	A&B	11.2	14.2		26.5	29.1	33.9							
19	5/ /17	A	15.3	20.6	23.2	29.6	34.3	37.2			13.9	18.9	24.2	29.5	31.6
20	5/29/17	A									18.5	24.7	28.3	31.2	32.8
Median . . . . .			6.8	15.2	19.2	27.3	30.9	33.2		7.6	13.5	19.6	25.1	28.4	30.
Woody's Standard Median			4.7	11.1	18.3	26.1	30.6	32.9		5.8	9.9	16.5	23.8	27.4	30.1

# BOSTON RESEARCH TESTS IN FRACTIONS

These tests are not generally available and are given in this connection to suggest to the teacher the possibility of becoming keen and active in the work of discovering pupils' errors. The Boston test in addition of fractions consists of six simple tests of four problems each, each test having a two-minute time limit. They cover the various types of problems in the addition of fractions, and they increase in difficulty from the first example in which the denominators are the same, up to the last in which the common denominator can be determined with difficulty by introspection. The test follows herewith:

## ADDITION OF FRACTIONS

Showing Examples Used in Tests in Addition of Fractions,  
December, 1915

Test 1. — Time, 2 minutes.

$$(1) \begin{array}{r} \frac{1}{4} \\ \frac{1}{4} \\ \hline \end{array}$$

$$(2) \begin{array}{r} \frac{3}{14} \\ \frac{1}{14} \\ \hline \end{array}$$

$$(3) \begin{array}{r} \frac{5}{18} \\ \frac{7}{18} \\ \hline \end{array}$$

$$(4) \begin{array}{r} \frac{1}{10} \\ \frac{7}{10} \\ \hline \end{array}$$

Test 2. — Time, 2 minutes.

$$(1) \begin{array}{r} \frac{1}{8} \\ \frac{1}{8} \\ \hline \end{array}$$

$$(2) \begin{array}{r} \frac{2}{7} \\ \frac{3}{14} \\ \hline \end{array}$$

$$(3) \begin{array}{r} \frac{2}{3} \\ \frac{1}{12} \\ \hline \end{array}$$

$$(4) \begin{array}{r} \frac{1}{3} \\ \frac{7}{15} \\ \hline \end{array}$$

Test 3. — Time, 2 minutes.

$$(1) \begin{array}{r} \frac{3}{8} \\ \frac{1}{8} \\ \hline \end{array}$$

$$(2) \begin{array}{r} \frac{5}{8} \\ \frac{1}{2} \\ \hline \end{array}$$

$$(3) \begin{array}{r} \frac{5}{7} \\ \frac{1}{14} \\ \hline \end{array}$$

$$(4) \begin{array}{r} \frac{14}{15} \\ \frac{2}{3} \\ \hline \end{array}$$

Test 4. — Time, 2 minutes.

$$(1) \begin{array}{r} \frac{1}{7} \\ \frac{9}{10} \\ \hline \end{array}$$

$$(2) \begin{array}{r} \frac{7}{8} \\ \frac{1}{4} \\ \hline \end{array}$$

$$(3) \begin{array}{r} \frac{3}{4} \\ \frac{3}{7} \\ \hline \end{array}$$

$$(4) \begin{array}{r} \frac{4}{8} \\ \frac{5}{8} \\ \hline \end{array}$$

Test 5. — Time, 2 minutes.

$$(1) \begin{array}{r} \frac{1}{10} \\ \frac{1}{8} \\ \hline \end{array}$$

$$(2) \begin{array}{r} \frac{4}{8} \\ \frac{5}{12} \\ \hline \end{array}$$

$$(3) \begin{array}{r} \frac{1}{8} \\ \frac{3}{8} \\ \hline \end{array}$$

$$(4) \begin{array}{r} \frac{1}{12} \\ \frac{1}{10} \\ \hline \end{array}$$

Test 6. — Time, 2 minutes.

$$(1) \begin{array}{r} \frac{1}{8} \\ \frac{9}{10} \\ \hline \end{array}$$

$$(2) \begin{array}{r} \frac{5}{8} \\ \frac{3}{8} \\ \hline \end{array}$$

$$(3) \begin{array}{r} \frac{1}{8} \\ \frac{9}{10} \\ \hline \end{array}$$

$$(4) \begin{array}{r} \frac{7}{12} \\ \frac{7}{10} \\ \hline \end{array}$$

The directions for scoring the test are not available, and without such directions definite comparison cannot be made. It seems worth while, however, to indicate the city medians for Boston, and these are summarized herewith in Table 25.



TABLE 25. — SUMMARY SHEET — CITY MEDIANS (BOSTON)

Addition of Fractions, December, 1915

GRADE	PUPILS TESTED	TEST 1		TEST 2		TEST 3		TEST 4		TEST 5		TEST 6	
		Speed Median (Number)	Accuracy Median (%)	Speed Median	Accuracy Median	Speed Median	Accuracy Median	Speed Median	Accuracy Median	Speed Median	Accuracy Median	Speed Median	Accuracy Median
VIII . .	1130	20.7	88.0	11.6	74.0	8.4	47.0	6.0	68.0	6.9	52.0	6.4	47.0
VII . . .	1243	16.5	87.0	10.1	73.0	7.3	46.0	5.3	69.0	6.3	55.0	5.7	48.0
VI . . .	1265	10.7	80.0	7.7	66.0	5.5	42.0	4.0	70.0	4.6	51.0	4.4	49.0

These tests as given in the Boston schools proved especially helpful in the work of analyzing the difficulties of pupils and devising drills to raise the efficiency of the children. This was evidenced by the increase in both speed and accuracy in tests given during the following spring to selected sixth grades. The teacher should find these tests in addition of fractions very useful, and she can make comparisons on the basis of her own rules for scoring.

The Boston tests in the subtraction of fractions will be equally suggestive and helpful to the teacher who is attempting to analyze the types of problems and the difficulties encountered by children in the solution of problems in subtraction of fractions. The tests as given follow herewith.

#### SUBTRACTION OF FRACTIONS

Showing Examples Used in Tests in Subtraction of Fractions,  
December, 1916

Test 1. — Time, 2 minutes.

(1)  $\frac{1}{4} - \frac{1}{4}$

(2)  $\frac{3}{4} - \frac{1}{4}$

(3)  $\frac{5}{8} - \frac{1}{8}$

(4)  $\frac{9}{16} - \frac{3}{16}$



Test 2. — Time, 2 minutes.

$$\begin{array}{llll} (1) \frac{1}{2} & (2) \frac{9}{7} & (3) \frac{8}{3} & (4) \frac{3}{4} \\ \hline \frac{1}{6} & \frac{3}{8} & \frac{3}{11} & \frac{4}{9} \end{array}$$

Test 3. — Time, 2 minutes.

$$\begin{array}{llll} (1) \frac{5}{8} & (2) \frac{3}{4} & (3) \frac{7}{5} & (4) \frac{7}{10} \\ \hline \frac{1}{10} & \frac{3}{10} & \frac{1}{12} & \frac{8}{15} \end{array}$$

Test 4. — Time, 2 minutes.

$$\begin{array}{llll} (1) 4 & (2) 6 & (3) 6 & (4) 6 \\ \hline 2\frac{1}{2} & 5\frac{3}{8} & 2\frac{3}{8} & 3\frac{1}{8} \end{array}$$

Test 5. — Time, 2 minutes.

$$\begin{array}{llll} (1) 9\frac{1}{2} & (2) 71\frac{3}{4} & (3) 71\frac{1}{2} & (4) 7\frac{1}{2} \\ \hline 1\frac{1}{2} & 0\frac{2}{7} & 4\frac{3}{8} & 2\frac{7}{15} \\ & & & \frac{11}{15} \end{array}$$

These tests increase in difficulty from the first to the fifth test and some of the examples in tests 4 and 5 are as difficult as any likely to appear in actual social and business practice. Instructions for the scoring of these tests are not at hand, but nevertheless the summary of the Boston medians is given herewith in Table 26.

TABLE 26. — SUMMARY SHEET — CITY MEDIANS (BOSTON)  
Subtraction of Fractions, December, 1916

GRADE	PUPILS	TEST 1		TEST 2		TEST 3		TEST 4		TEST 5	
		Speed Median (Number)	Accuracy Median %	Speed Median	Accuracy Median	Speed Median	Accuracy Median	Speed Median	Accuracy Median	Speed Median	Accuracy Median
VIII . . . .	1239	22.5	91.0	7.3	86.0	6.1	65.0	18.0	99.0	6.4	81.0
VII . . . .	1283	19.7	84.0	6.0	85.0	5.6	61.0	14.2	97.0	5.2	66.0
VI . . . .	1499	15.1	73.0	4.9	76.0	4.6	51.0	11.9	85.0	4.6	64.0

Tests were also devised in the multiplication and division of fractions. Some of these tests are quite difficult and yet for diagnostic purposes they will show the ability of children to multiply or divide fractions and mixed numbers. The tests are indicated herewith, and city medians are summarized in Table 27.

MULTIPLICATION AND DIVISION OF FRACTIONS

Showing Examples Used in Tests in Multiplication and Division of Fractions, December, 1917

Multiplication of Fractions. — Test 1. Time, 2 minutes.

(1)  $\frac{1}{8} \times 6$       (2)  $\frac{7}{9} \times 8$       (3)  $\frac{5}{6} \times 12$       (4)  $12 \times \frac{5}{18}$

Multiplication of Fractions. — Test 2. Time, 4 minutes.

(1)  $246\frac{1}{5}$       (2)  $573\frac{4}{5}$       (3)  $275\frac{3}{4}$       (4)  $456\frac{1}{3}$       (5)  $189\frac{1}{2}$

$$\begin{array}{r} 5 \\ \hline \end{array} \quad \begin{array}{r} 5 \\ \hline \end{array} \quad \begin{array}{r} 8\frac{3}{4} \\ \hline \end{array} \quad \begin{array}{r} 2 \\ \hline \end{array} \quad \begin{array}{r} 5\frac{1}{2} \\ \hline \end{array}$$

Multiplication of Fractions. — Test 3. Time, 2 minutes.

(1)  $4\frac{7}{8} \times \frac{1}{8}$       (2)  $7\frac{1}{2} \times \frac{2}{3}$       (3)  $5\frac{1}{2} \times \frac{3}{4}$       (4)  $\frac{5}{8} \times 2\frac{2}{3}$

Multiplication of Fractions. — Test 4. Time, 5 minutes.

(1)  $32\frac{1}{3}$       (2)  $84\frac{1}{3}$       (3)  $29\frac{3}{4}$       (4)  $25\frac{3}{4}$       (5)  $19\frac{1}{8}$

$$\begin{array}{r} 69\frac{1}{2} \\ \hline \end{array} \quad \begin{array}{r} 79\frac{1}{5} \\ \hline \end{array} \quad \begin{array}{r} 28\frac{1}{3} \\ \hline \end{array} \quad \begin{array}{r} 17\frac{2}{3} \\ \hline \end{array} \quad \begin{array}{r} 97\frac{1}{2} \\ \hline \end{array}$$

Division of Fractions. — Test 5. Time, 2 minutes.

(1)  $\frac{3}{4} \div 8$       (2)  $9 \div \frac{3}{8}$       (3)  $6 \div \frac{4}{5}$       (4)  $8 \div \frac{3}{5}$

Division of Fractions. — Test 6. Time, 4 minutes.

(1)  $5678\frac{1}{3} \div 5$       (2)  $2789\frac{2}{3} \div 4$       (3)  $2467 \div 8\frac{1}{4}$       (4)  $6752 \div 12\frac{1}{2}$

Division of Fractions. — Test 7. Time, 3 minutes.

(1)  $\frac{3}{5} \div \frac{1}{3}$       (2)  $3\frac{3}{4} \div \frac{1}{5}$       (3)  $5\frac{4}{5} \div \frac{2}{3}$       (4)  $6\frac{2}{5} \div \frac{4}{5}$

It will be observed, from Table 27, that the scores in tests 2, 4, and 6 are very low, indicating that they were not well chosen. Referring to these particular tests, Dr. Ballou of the Boston bureau says:

"It is probably true that there is no great use for the type of work shown in these three tests in practical life, but the business world does require it to some extent; business courses in our high schools require the processes, and the new course of study requires this work. In view of these three conditions, it was thought best to include these three tests in order that we might have some facts on which to base the development of our work in multiplication and division of fractions."

TABLE 27. — SUMMARY SHEET — CITY MEDIANS (BOSTON)

## Multiplication and Division of Fractions

GRADE	PUPILS TESTED	MULTIPLICATION								DIVISION							
		Test 1		Test 2		Test 3		Test 4		Test 5		Test 6		Test 7			
		Speed Median (Number)	Accuracy Median %	Speed Median	Accuracy Median	Speed Median	Accuracy Median	Speed Median	Accuracy Median	Speed Median	Accuracy Median	Speed Median	Accuracy Median	Speed Median	Accuracy Median		
VIII	1027	11.1	93	8.8	63	7.6	85	4.7	0	10.1	75	3.3	29	10.3	79		
VII	1196	8.4	88	7.7	38	6.4	81	4.2	0	8.2	59	2.9	0	8.5	68		
VI	1290	6.2	13	8.2	0	4.7	0	5.6	0	5.4	0	3.2	0	4.9	0		

On the basis of this statement we may expect that the course of study in Boston will be improved by much elimination and better adaptation to business usage. Tests will be used more and more in the work of revising courses of study.

**Cleveland Survey Tests.** — One of a number of hopeful signs in the development of arithmetic tests is the clear recognition that they should be of direct value in helping the children. This recognition is leading quite clearly to the more extensive development of diagnostic tests. The tests

used in the Cleveland survey were prepared in coöperation with Mr. Courtis, who recognized as clearly as any one else the need of supplementary work in order to make his standard tests, series A and B, of sufficient value to the teacher whose duty is to improve the pupils in their work. No attempt will be made in this place to describe or discuss fully the Cleveland survey tests.<sup>1</sup> The tests, now slightly revised, are composed of 15 different sets of examples designated as A, B, C, D, E, F, G, H, I, J, K, L, M, N, O. They are intended to cover the "fundamentals" of arithmetic. Of the 15 tests, four are in addition; two in subtraction; three in multiplication; four in division; and two in fractions. They constitute to an extent a spiral arrangement of tests, increasing in difficulty from A to O. The actual time covered by the tests is 22 minutes, and this combined with the time necessary to pass from one test to another led to the direction, in connection with the Grand Rapids survey, that two days be taken for the tests; the first nine sets being given on the first day and the remaining six sets being given on the following day. As indicated, the sets were devised in coöperation with Mr. Courtis and they follow the Courtis Practice Forms more or less closely. These forms were used in the Grand Rapids schools so that the results secured in Grand Rapids may be considered quite satisfactory. The results in the Cleveland schools were more satisfactory in the lower grades but a little less so in the upper grades. Table 28, following herewith, shows the average of the median scores in each of the arithmetic tests for grades 3 to 8 in Cleveland and Grand Rapids. This table may be considered as setting tentative standards for the Cleveland survey tests for the various grades.

<sup>1</sup> For further discussion see Judd, C. H., "Measuring the Work of the Public School," a volume of *The Cleveland Survey; School Survey of Grand Rapids, Mich.*, Chap. VI; "Arithmetic Tests and Studies in the Psychology of Arithmetic," by Counts, G. S.; Supplementary Educational Monograph, whole number IV.

TABLE 28. — AVERAGES OF MEDIAN SCORES IN EACH ARITHMETIC TEST FOR GRADES 3 TO 8, CLEVELAND AND GRAND RAPIDS COMBINED

SET	GRADE					
	3	4	5	6	7	8
A . . . . .	13.4	17.1	21.9	24.9	27.0	28.9
B . . . . .	8.9	12.8	16.6	19.5	21.1	25.8
C . . . . .	6.5	11.7	14.8	16.8	18.2	19.9
D . . . . .	6.3	11.4	15.0	17.7	20.3	22.8
E . . . . .	4.3	5.0	5.9	6.7	7.4	8.0
F . . . . .	2.0	4.5	6.6	7.7	9.1	10.6
G . . . . .	2.0	3.6	5.1	5.5	6.0	6.7
H . . . . .			5.6	6.0	7.7	8.6
I . . . . .	0.6	1.0	1.7	3.1	4.0	4.7
J . . . . .	1.9	3.0	3.9	4.4	5.1	6.1
K . . . . .		4.0	5.6	7.0	9.4	11.4
L . . . . .		1.7	2.7	3.2	3.8	4.4
M . . . . .	1.4	2.4	3.4	4.1	4.7	5.4
N . . . . .		0.8	1.1	1.6	1.9	2.4
O . . . . .				3.3	4.3	5.2

The four addition sets, A, E, J, M, follow herewith, and they may be considered as representative of the spiral arrangement and diagnostic character of the Cleveland tests. It will be observed that the examples increase in difficulty and

## Set A — Addition.

1	6	9	0	4	1	7	9	3	2	1	3	6
2	6	5	1	2	3	7	6	0	4	5	8	9
0	3	8	9	7	8	2	1	4	8	0	2	2
7	2	1	9	6	0	5	6	7	9	5	7	1
4	7	0	3	1	2	5	6	7	5	8	6	9
6	8	8	5	4	9	8	0	2	1	3	5	0
4	2	9	7	4	5	7	4	8	0	3	9	2
3	2	3	8	0	2	1	9	6	0	4	1	8

5	0	6	2	4	5	1	6	3	7	9	0	4
<u>7</u>	<u>4</u>	<u>3</u>	<u>1</u>	<u>8</u>	<u>9</u>	<u>0</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>8</u>	<u>6</u>	<u>5</u>

Set E — Addition.

5	2	9	2	6	1	4	9
2	8	8	8	3	4	6	7
2	8	0	5	4	2	5	1
0	5	7	0	8	5	3	5
<u>4</u>	<u>1</u>	<u>6</u>	<u>6</u>	<u>8</u>	<u>4</u>	<u>4</u>	<u>3</u>
6	2	6	8	5	4	1	3
7	7	2	5	9	0	4	7
8	3	3	1	6	8	1	2
5	4	9	3	3	5	8	9
<u>5</u>	<u>1</u>	<u>3</u>	<u>8</u>	<u>8</u>	<u>5</u>	<u>4</u>	<u>6</u>

Set J — Addition.

7	9	4	7	2	9	6	7	7	8	9	4	3	2
5	2	5	1	9	6	9	1	8	0	5	3	1	1
4	4	8	9	4	2	6	5	5	7	3	7	7	6
2	8	1	4	8	4	7	1	4	1	4	7	6	6
6	2	4	3	5	7	0	4	1	8	6	0	9	1
0	7	8	2	1	1	4	6	8	5	2	2	6	8
5	5	5	8	5	3	3	5	2	1	3	9	3	6
1	3	1	5	2	9	7	3	1	3	9	5	4	9
8	6	3	2	4	2	1	3	3	7	2	6	5	7
3	1	9	7	3	3	6	7	9	4	2	3	4	5
2	4	6	7	6	8	0	6	8	9	8	4	2	2
9	8	3	1	7	5	6	1	4	4	5	8	9	2
<u>9</u>	<u>8</u>	<u>5</u>	<u>9</u>	<u>6</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>5</u>	<u>4</u>	<u>6</u>	<u>8</u>	<u>9</u>	<u>4</u>

Set M — Addition.

7493	8937	8625	2123	5142	3691
9016	6345	4091	1679	0376	4526
6487	2783	3844	5555	4955	7479
7591	4883	8697	6331	9314	2087
<u>6166</u>	<u>1341</u>	<u>7314</u>	<u>6808</u>	<u>5507</u>	<u>8165</u>



5226	9149	6268	9397	7337	8243
2883	8467	7725	6158	2674	6429
2584	0251	8331	3732	9669	9298
0058	7535	5493	4641	5114	7404
2308	<u>5223</u>	<u>3918</u>	<u>7919</u>	<u>8154</u>	<u>2575</u>

lend themselves quite well to diagnostic purposes. Set A tests the pupils' knowledge of the addition combinations; set E is a simple test in column addition; set J involves more difficult column addition; and set M requires carrying as well as column addition, and conforms to business usage more closely than the Curtis Series B.

Subtraction is tested in sets B and F; multiplication in sets C, G, and L; division in sets D, I, K, and N; and fractions in sets H and O. In each of the fundamental processes and in fractions, the first set is quite simple and each later set grows more difficult. The detail shown above in addition represents the plan in each process. The diagnostic use of the Cleveland tests may best be illustrated by taking some actual cases. Mary S., a pupil in the eighth grade, thirteen years old, makes no mistakes in any of the tests in addition. She solves every problem attempted, but her score for the various tests in addition is low. They run as follows: A 19, E 5, J 4, M 4. The standards for these tests in the eighth grade as shown by the Cleveland and Grand Rapids reports (see Table 28) are A 28.0, E 8.0, J 6.1, M 5.4. It appears, therefore, that Mary S. needs to continue her present accuracy in addition but increase her speed. In subtraction she makes a mistake in test B in subtracting 9 from 11, but her score in subtraction is low in both test B and test F. In multiplication she is inaccurate as well as slow, making two mistakes in set C which are the simple combinations. In division she is exceedingly slow. In set K in division the standard score for the eighth grade according to Table 28 is 11.4. Mary S. has a score of 1.

Turning to the record of Hazel R., a thirteen-year-old girl

in the eighth grade, it appears that her scores are down throughout, but that her inaccuracies are chiefly in sets H and O. These tests, as will be noted above, relate to simple fractions. However, in addition, Hazel R. has difficulty. She solves twenty problems in set A all correctly and four in set E correctly, but in set J involving a column of thirteen figures she fails on every problem attempted. In set M, five addends of four-place numbers, she fails on one out of three attempted.

It is evident, from the above, that rather detailed analysis of the pupils' difficulties is easily made from the results of the Cleveland tests. When pupils pass set A with proper speed and accuracy it means that they know the addition combinations. When they fail on set J it means that the more complex numbers involve too much mental effort or that the drill on decades has not been sufficient, doubtless the latter, because many pupils who know that four and eight are twelve fail when the combination is twenty-four and eight. In like manner, a pupil's paper will show for the other fundamental processes and simple fractions just where his difficulties begin and, therefore, just where the teacher needs to begin in order to give the necessary help. How to analyze the arithmetic difficulties in an entire city system through the use of the Cleveland survey tests has been shown by Dr. George S. Counts in the *School Review Educational Monograph*, number IV.

**Kansas Diagnostic Tests in Arithmetic.** — The Bureau of Educational Measurement and Standards, of the State Normal School at Emporia, Kansas, has done a notable work in connection with the measurement of subject matter. The fact that the tests in arithmetic, finally put out by this Bureau, were diagnostic in nature, may be taken as a further indication that, in the future, increasing importance will be attached to diagnostic tests. The Kansas Tests in Arithmetic have not been extensively tried out and so are not well

standardized. Tentative standards, however, have been issued covering the 21 tests. These are the midyear scores and are based upon results of testing from 300 to 1200 pupils in each of the grades. Table 29, which follows herewith, gives these tentative standards for the first six tests.

TABLE 29. — TENTATIVE STANDARDS FOR THE KANSAS DIAGNOSTIC TESTS IN ARITHMETIC

The number of pupils taking the tests varies from about 300 to over 1200. Midyear scores. R — Rate or number of examples done. A — Accuracy or per cent of examples correct

GRADE	IV		V		VI		VII		VIII	
Test No.	R	A	R	A	R	A	R	A	R	A
1	7.8	100	12.3	100	10.1	100	12.1	100	11.9	100
2	3.7	60	7.3	100	7.0	100	8.0	100	8.9	100
3	3.1	57	4.9	75	5.1	79	5.5	83	6.2	84
4	2.1	40	2.9	60	3.4	68	4.3	79	4.6	88
5	4.2	52	5.2	64	5.3	63	5.4	63	6.1	66
6	1.9	38	3.4	70	3.2	74	4.7	70	4.5	100

These six tests cover simple work in addition, subtraction, multiplication, and division and are reproduced with the consent of the author.

### PART I — TESTS 1-6

#### OPERATIONS WITH INTEGERS

At.....

Test 1 — Addition.

Rt.....

4	5	2	0	1	7	6	7	3	2	3	9
7	5	6	3	1	2	8	7	8	4	3	4
2	9	7	8	4	3	4	0	9	0	6	5
8	8	5	4	4	1	0	0	7	6	6	3
0	9	9	6	5	5	2	1	1	8	7	7
5	2	1	1	8	7	7	4	3	3	0	9

Test 2 — Subtraction.

At.....

Rt.....

$\begin{array}{r} 37 \\ 5 \end{array}$	$\begin{array}{r} 94 \\ 8 \end{array}$	$\begin{array}{r} 60 \\ 3 \end{array}$	$\begin{array}{r} 27 \\ 6 \end{array}$	$\begin{array}{r} 39 \\ 7 \end{array}$	$\begin{array}{r} 41 \\ 8 \end{array}$	$\begin{array}{r} 77 \\ 3 \end{array}$	$\begin{array}{r} 53 \\ 9 \end{array}$
$\begin{array}{r} 65 \\ 2 \end{array}$	$\begin{array}{r} 80 \\ 4 \end{array}$	$\begin{array}{r} 92 \\ 5 \end{array}$	$\begin{array}{r} 70 \\ 3 \end{array}$	$\begin{array}{r} 68 \\ 2 \end{array}$	$\begin{array}{r} 58 \\ 9 \end{array}$	$\begin{array}{r} 26 \\ 9 \end{array}$	$\begin{array}{r} 43 \\ 8 \end{array}$
$\begin{array}{r} 95 \\ 4 \end{array}$	$\begin{array}{r} 50 \\ 7 \end{array}$	$\begin{array}{r} 36 \\ 1 \end{array}$	$\begin{array}{r} 34 \\ 8 \end{array}$	$\begin{array}{r} 44 \\ 6 \end{array}$	$\begin{array}{r} 25 \\ 3 \end{array}$	$\begin{array}{r} 63 \\ 7 \end{array}$	$\begin{array}{r} 57 \\ 9 \end{array}$

At.....

Test 3 — Multiplication.

Rt.....

$\begin{array}{r} 6572 \\ 6 \end{array}$	$\begin{array}{r} 6750 \\ 9 \end{array}$	$\begin{array}{r} 5863 \\ 2 \end{array}$	$\begin{array}{r} 3754 \\ 5 \end{array}$	$\begin{array}{r} 2845 \\ 8 \end{array}$
$\begin{array}{r} 4936 \\ 4 \end{array}$	$\begin{array}{r} 9327 \\ 7 \end{array}$	$\begin{array}{r} 8274 \\ 3 \end{array}$	$\begin{array}{r} 8409 \\ 6 \end{array}$	$\begin{array}{r} 6391 \\ 9 \end{array}$
$\begin{array}{r} 5482 \\ 2 \end{array}$	$\begin{array}{r} 8609 \\ 5 \end{array}$	$\begin{array}{r} 3679 \\ 8 \end{array}$	$\begin{array}{r} 2758 \\ 4 \end{array}$	$\begin{array}{r} 4658 \\ 7 \end{array}$
$\begin{array}{r} 9653 \\ 3 \end{array}$	$\begin{array}{r} 3174 \\ 6 \end{array}$	$\begin{array}{r} 2874 \\ 9 \end{array}$	$\begin{array}{r} 7901 \\ 2 \end{array}$	$\begin{array}{r} 2179 \\ 5 \end{array}$

At.....

Test 4 — Division.

Rt.....

$8\overline{)3840}$	$4\overline{)7432}$	$7\overline{)3534}$	$3\overline{)9430}$	$6\overline{)4680}$
$9\overline{)8577}$	$2\overline{)6370}$	$5\overline{)9310}$	$8\overline{)7512}$	$4\overline{)3820}$
$7\overline{)9653}$	$3\overline{)5781}$	$6\overline{)6720}$	$9\overline{)5373}$	$2\overline{)5130}$

## Test 5 — Addition.

7862	6809	8941	5917	6772	7864	1249
5013	7623	7910	4814	6028	7883	8975
1761	5299	9845	9007	6535	8240	9005
5872	6601	8522	6975	2340	8969	1573
<u>3739</u>	<u>3496</u>	<u>1046</u>	<u>1227</u>	<u>2319</u>	<u>6794</u>	<u>3203</u>
8758	2462	1247	4319	6794	3293	7917
2350	9869	3573	2358	5420	7805	4304
3197	4572	1081	5795	4570	7642	9027
2338	6420	7805	4314	8028	7803	9975
<u>5917</u>	<u>6772</u>	<u>9864</u>	<u>1249</u>	<u>8758</u>	<u>2462</u>	<u>1247</u>

At.....

Rt.....

## Test 6 — Division.

82) <u>3854</u>	43) <u>1591</u>	74) <u>2664</u>	31) <u>1953</u>
63) <u>3591</u>	94) <u>4042</u>	21) <u>1407</u>	53) <u>4452</u>
83) <u>5312</u>	42) <u>672</u>	71) <u>5183</u>	32) <u>2304</u>
62) <u>2108</u>	93) <u>5022</u>	23) <u>874</u>	51) <u>2703</u>
84) <u>7140</u>	41) <u>3567</u>	73) <u>6278</u>	33) <u>1386</u>
64) <u>4992</u>	92) <u>6624</u>	24) <u>1008</u>	52) <u>3484</u>

At.....

Rt.....

The value of these tests can be ascertained only by extended use. They follow rather closely the line of testing marked out by the Cleveland survey tests. Tests 7 to 11 cover operations with integers in the fundamentals and are similar to tests 1 to 6, except that they are more difficult. Tests 12 to 16 cover operations with common fractions. These tests are simple and not beyond reasonable business demands. They should be of value in helping teachers to locate a pupil's difficulties in handling fractions. Tests 17 to 21 deal with

the multiplication and division of decimals and relate particularly to the problem of placing the decimal point. There will quite surely be objections to these tests because the examples in the tests are more difficult than problems in decimals which appear in common business practice. It is apparent that the test in decimals has been based upon an analysis of textbook material rather than actual usage.

**The Teacher's Problem.** — In view of the many tests in arithmetic which are now available, the teacher may feel confused and uncertain as to just how she should proceed. Apparently the best method would be to master, one at a time, the details of testing achievement in arithmetic. The teacher will do well to start with the tests which developed first historically, namely, the Courtis tests, series B. While these tests have defects which are admitted even by their author, yet they are better standardized than any other tests at the present time and so will best answer the purpose of measuring achievement in arithmetic, particularly for upper grade pupils. They are as serviceable as any other tests at the present time. When the teacher feels confident that she understands the Courtis tests, knows how to administer them, to grade them and to apply them to the remedial work in her own schoolroom, then she may properly take up the Woody scales for the fundamentals. These are quite well standardized. They are easily administered and they are valuable for diagnostic purposes. Many teachers will not use other tests. Some however, will begin to see the advantage of standard tests and will desire to take up the Cleveland survey tests, applying them to their own schools, comparing their schools with others which have been measured by these tests, but particularly using them for locating the weaknesses of their own pupils. These tests were originally designed for diagnostic purposes and they seem well calculated to serve this purpose.

In a year of earnest work, the teacher should master testing pupils in abstract numbers, should locate the shortcomings



of her class, and should make decided progress toward bringing the entire class to standard. Many teachers will prefer to wait until a second year to begin the use of the reasoning tests. This may be a wise procedure because their use should be followed by a thorough study of reasoning problems, the progressive steps which pupils must take in mastering such problems, and the business demands for such problems. A valuable study on reasoning problems, and the successive steps in their mastery, is found in the Connersville Course of Study in Elementary Mathematics. Since figuring which is done in actual business is always connected with situations requiring reasoning, the teacher should not fail to carry forward her work until she has mastered the details of administering, scoring, and interpreting reasoning tests.

**The Next Step.** — What is the next step in measurement in arithmetic? Some say it is to devise tests for the measurement of the higher processes in arithmetic. This may be so, but it is to be hoped that before such tests are formulated, the needs of common business practice will be more fully determined. If tests were now formulated for denominate numbers, percentage and its applications, mensuration, etc., they would almost surely represent merely textbook and schoolroom viewpoints. The results would doubtless be less satisfactory than the Kansas tests in decimals. It is to be hoped, therefore, that the more fundamental work of determining the actual community and business demands of Arithmetic will be carried much further before any attempt is made to extend measurement in arithmetic to the higher processes. Progress is being made along this line<sup>1</sup> and in time we may hope to have a type of arithmetic throughout

<sup>1</sup> See particularly, Wilson, G. M., "The Social and Business Usage of Arithmetic," Teachers College Bureau of Publications; Mitchell, H. Edwin, "Some Social Demands in the Course of Study in Arithmetic," Teachers College Bureau of Publications; the 1st, 2d, and 3d reports of the Committee on Minimum Essentials of the National Education Association, Public School Publishing Co.; and the Iowa Elimination Reports.

the entire course, which is directly applicable to business usage and which is so taught as to further the intelligent use of arithmetic in business. In the meantime, teachers are quite safe in furthering the work of measurement in arithmetic in the fundamental processes and simple fractions. Teachers may assume that mastery here is essential, and that measurement is valid so long as applied only to the formal aspects of the subject.

# BIBLIOGRAPHY

For the convenience of the reader, the references are listed under the tests referred to in the text. The references are numbered in order by the arabic system.

## I. COURTIS STANDARD RESEARCH TESTS, SERIES B.

The tests may be purchased from S. A. Courtis, 82 Eliot St., Detroit, Michigan. In writing state the number of pupils to be tested.

1. Courtis, S. A., "Annual Accounting," 1913-1916; Department of Coöperative Research, 82 Eliot St., Detroit.
2. Courtis, S. A., "Measurement of Growth and Efficiency in Arithmetic," *Elementary School Teacher*, Vol. 10, pp. 55-74; Vol. 11, pp. 171-185; pp. 360-370; pp. 528-539; Vol. 12, pp. 127-137.
3. Courtis, S. A., "Educational Diagnosis," *Educational Administration and Supervision*, February, 1915.
4. Courtis, S. A., "Courtis Tests in Arithmetic: Value to Superintendents and Teachers," Fifteenth Yearbook of the National Society for the Study of Education, Part I, pp. 91-106.
5. Buckingham, B. R., "The Courtis Tests in the Schools of New York," *Journal of Educational Psychology*, April, 1914.
6. Haggerty, M. E., "Arithmetic — A Coöperative Study in Educational Measurements," Bulletin No. 27, Indiana University.
7. Haggerty, M. E., "Studies in Arithmetic," Bulletin No. 32, Indiana University.
8. Monroe, W. S., "The Courtis Standard Tests in Arithmetic in Twenty-four cities," Bureau of Educational Measurements and Standards, Bulletin No. 4, Emporia, Kansas.
9. Ballou, Frank W., In bulletins of the Department of Educational Investigation and Measurement of Boston. No. X: "The Courtis Standard Tests in Boston, 1912-1915"; No. XI: "Provisional Standards in Arithmetic"; No. XIII: "Value to Teacher,

Principal, and Superintendent of Individual and Class Records from Standard Tests."

10. School Surveys: Butte, Montana; Salt Lake City; Leavenworth, Kansas; New York City.

## II. WOODY SCALES IN THE FUNDAMENTALS

Supplies available through Teachers College Bureau of Publications, West 120th Street, New York City.

11. Woody, Clifford, "Measurement of Some Achievements in Arithmetic," Teachers College Bureau of Publications.
12. Monroe, W. S., "An Experimental and Analytical Study of Woody's Arithmetic Scales," Series B, *School and Society*, VI, pp. 412-420, October 6, 1917.
13. Anderson, C. J., "The Use of the Woody Scale for Diagnostic Purposes," *Elementary School Journal*, June, 1918, pp. 770-781.

## III. BOSTON RESEARCH TESTS IN FRACTIONS

Copies not available for distribution.

14. Ballou, Frank W., In bulletins of the Department of Educational Investigation and Measurement, of Boston. No. VII: "Arithmetic. Research Tests in Addition of Fractions"; No. XV: "Arithmetic. Achievement of Pupils in Common Fractions."

## IV. THE CLEVELAND SURVEY TESTS IN ARITHMETIC

Supplies may be secured by addressing Director Charles H. Judd, of the School of Education, Chicago University.

15. Judd, Charles H., "Measuring the Work of the Public Schools." A volume in the Cleveland Survey Series.
16. Counts, George S., "Arithmetic Tests and Studies in the Psychology of Arithmetic," Supplementary Educational Monograph, Whole Number IV, Chicago University Press.
17. O'Hern, Joseph P., "Practical Application of Standard Tests," *Elementary School Journal*, May, 1918, pp. 662-679.
18. Smith, J. H., "Individual Variation in Arithmetic," *Elementary School Journal*, November, 1916, pp. 195-200.
19. Heckert, J. W., "The Cleveland Survey Tests in Arithmetic in the Miami Valley," *Elementary School Journal*, Vol. 17, February, 1918, p. 447.
20. School Surveys: Cleveland, Grand Rapids, St. Louis.

V. KANSAS DIAGNOSTIC TESTS IN ARITHMETIC

Supplies may be secured from the Bureau of Educational Measurements and Standards, Emporia, Kansas.

21. Monroe, W. S., "Ability to Place the Decimal Point," *Elementary School Journal*, 18: 287-293, December, 1917.
22. Wilson, G. M., "The Proper Content of a Standard Test," *Elementary School Journal*, 19: 375-381, January, 1919.
23. Monroe, W. S., "Diagnostic Tests in Arithmetic," *Elementary School Journal*, 19: 585-607, April, 1919.

VI. STONE REASONING TESTS

24. Stone, C. W., "Arithmetical Abilities and Some Factors Determining Them," Teachers College Contributions to Education, No. 19, Columbia University, New York City.
25. Stone, C. W., "Standardized Reasoning Tests in Arithmetic and How to Utilize Them," Teachers College Contributions to Education, No. 83.
26. School Surveys: Butte; Salt Lake City; Nassau County, New York.

VII. BETTER SELECTION OF SUBJECT MATTER IN ARITHMETIC

27. McMurry, Frank M., "The Elimination of Useless Material," 1904 Yearbook, National Education Association.
28. Wilson, G. M., "Course of Study in Elementary Mathematics," Warrick and York, Publishers, Baltimore, Maryland. (In press.)
29. Wilson, G. M., "A Survey of the Social and Business Usage of Arithmetic," Sixteenth Yearbook of National Society for the Study of Education, Part I, pp. 128-142.
30. Wilson, G. M., "A Survey of the Social and Business Usage of Arithmetic," Dissertation, Teachers College Bureau of Publications, Columbia University, New York City.
31. First and Second Iowa Elimination Reports.
32. Mitchell, H. Edwin, "Some Social Demands on the Course of Study in Arithmetic," Seventeenth Yearbook of the National Society for the Study of Education, Part I, pp. 7-17.
33. Wise, Carl T., "A Survey of Arithmetic Problems Arising in Various Occupations," *Elementary School Journal*, 20: 118-136, October, 1919.

## CHAPTER V

### THE MEASUREMENT OF READING

"Most of the reading done by people after leaving school will be silent reading. I am of the opinion that much more stress should be laid upon it in the schoolroom. The Monroe Standardized Silent Reading Tests are a help to the teacher at the beginning of the school year in finding just where the pupils stand in reading. They also help her to discover the individual needs of each pupil. Tests given at the close of the term show the progress made." This statement was made by a teacher of a third grade after she had used the Monroe Standardized Silent Reading Tests for one half year to determine the ability and the progress of her children in reading. It reveals two or three important considerations for successful reading achievement.

Teachers are beginning to realize the importance of placing more emphasis upon silent reading, since so much reading outside of the schoolroom is silent and since success in a vocation and happiness in leisure time depend upon the individual's ability to grasp the thought on the written page. The statement also reveals to the teacher the importance of knowing in a definite and objective manner what her class as a whole and what each individual in her class can do in reading. It is also necessary to know in definite and objective terms the amount of progress made by the class as a whole and by each individual student at the expiration of a certain period of time.

The importance of silent reading cannot be overestimated. The ability to grasp quickly the thought in a paragraph, a chapter, or a book, is one of the biggest contributing factors



to the success of any individual. The ability to master quickly the thought in modern literature, current events, history, etc., determines to a large extent the progress and development of the individual among his fellow men. It has much to do with his happiness during his whole life. If, therefore, reading plays such a significant part in life, is it not a matter of considerable importance that, in the training of the child to master the symbols with which ideas are expressed, the best methods of instruction be used, his capacity be trained to the fullest extent with the least degree of waste, and his appreciation of the best that is said and written be developed? To this end the reading test as has been indicated is a most helpful instrument.

Among the different factors that contribute to good reading, which is made up of the ability to interpret and remember, there are two which are of great importance and which the teacher can determine in a definite manner. These factors are, first, the power of comprehending thought, and second, the rate of reading. These two factors are important in both kinds of reading, namely, silent and oral.

Since it is not the purpose of this work to make an exhaustive study of all the tests on any one subject, a treatment is given of those reading tests only which, on account of their aim, simplicity of application, and extent of use are most serviceable in the hands of teachers. The following tests meet this purpose: Thorndike's Scale Alpha 2, Curtis's Silent Reading Test No. 2, Monroe's Standardized Silent Reading Tests, Haggerty's Visual Vocabulary Test, and Gray's Oral Reading Tests.

#### MONROE'S STANDARDIZED SILENT READING TESTS

**Aim.** — The aim of these tests is to determine the rate at which children read and the extent to which they are able to comprehend the thought in written discourse. This is accomplished by having them read silently a given sentence



or paragraph and then write answers to questions on the subject matter they have read.

**Description of Test.** — Monroe's Standardized Silent Reading Tests have been constructed from sentences taken "from school readers and other books which children read." They are intended for grades three to twelve inclusive. They consist of three tests:

Test one for Grades 3, 4, and 5.

Test two for Grades 6, 7, and 8.

Test three for Grades 9, 10, 11, and 12.

Tests one and two have each three forms, forms 1, 2, and 3; test three has two forms, forms 1 and 2. Each form is of the same degree of difficulty, but has different subject matter so that the same class can be examined three times without using the same information. The first six paragraphs of test 1, form 1, which has in all sixteen paragraphs, are given in order to show the nature of the tests:

Rate  
Value 8

No. 1

"I am not playing, little girl," said the squirrel. "I am running to my home in the hollow tree. Don't you hear my babies calling me? I must feed them."

Where was the home of the squirrel?

In the . . . . .

Compre-  
hension  
Value 1.3

Rate  
Value 7

No. 2

The little Pilgrim girls carried their work boxes to the dame-schools and learned to sew and knit as well as to read and write.

Where did the girls go with their work boxes?

To the . . . . .

Compre-  
hension  
Value 1.3

Rate Value 7	<p>No. 3</p> <p>When the white men first came to this country they found the red men, or Indians, living in wigwams, made of long poles and covered with skins.</p> <p>Which people lived here first, the white or red?</p> <p>.....</p>	Compre- hension Value 1.3
Rate Value 9	<p>No. 4</p> <p>Hiawatha was a little Indian boy. He had no father and no mother. He lived with his grandmother, Nokomis. His home was in a wigwam. Draw a line under the word that tells whom Hiawatha lived with.</p> <p>Father, aunt, mother, uncle, sister, grandmother.</p>	Compre- hension Value 1.4
Rate Value 6	<p>No. 5</p> <p>The cabin of Uncle Tom was a small log building close adjoining to "the house," as the negro designates his master's dwelling.</p> <p>Of what material was Uncle Tom's cabin built?</p> <p>.....</p>	Compre- hension Value 1.4
Rate Value 7	<p>No. 6</p> <p>A crab who lived in a sand-hill was sitting at his door in the sun eating a rice cake. An ape went by, carrying an orange seed.</p> <p>Where did the crab live?.....</p>	Compre- hension Value 1.5

Each test begins with a simple exercise and increases in comprehension value with almost each succeeding exercise. The measure of a child's ability to understand or comprehend what he reads in each exercise forms the comprehension value and is placed in the right-hand margin opposite each

exercise. The sum of the values of those exercises done correctly in five minutes forms his comprehension score.

It is important not only to know the extent to which a child can grasp the thought in an exercise, but also the time in which it takes him to grasp this thought. The child who can grasp the thought in an exercise in three minutes has greater reading ability than the child who requires five minutes to grasp the thought in the same exercise. Each exercise is, therefore, given a rate value which represents the number of words read per minute in careful reading. This value is placed on the left-hand margin opposite each exercise. The pupil's reading score is the sum of the rate values given to the different exercises which he reads in five minutes.

**Giving the Test.** — The application of this reading test is very simple. Full instructions are given on the front page of each test. Each child is provided with a separate test. After he has written his name, age, grade, etc., he is given a preliminary test which explains how he is to proceed with the regular test. The value of the results will depend on the accuracy with which these simple instructions are observed.

**Scoring the Results.** — A class record sheet is provided with each test. A copy of this record sheet with the results from a 3-A grade is given in Table 30.

The score for the class of thirty-nine children reported in Table 30 is given in terms of a median instead of an average. This median score is the score on the middle paper in the group of papers that is being recorded. In March, thirty-nine children were tested. The comprehension score on the 20th paper, counted from the paper receiving the lowest score, was 10.6. In a similar manner the median score for the rate of reading was determined.

**Interpreting and Using Results.** — After a class has been tested and the results tabulated, the important work of interpreting these results and applying them to particular needs still remains. Unless these two steps are developed to the

TABLE 30. — THE RESULTS FROM MONROE'S STANDARDIZED SILENT READING TESTS GIVEN TO A 3-A GRADE IN MARCH 1918.

CLASS RECORD SHEET

City D  
Teacher N. E.

School B

Grade 3-A

Date March 22, 1918.

RATE SCORE		COMPREHENSION SCORE		INSTRUCTIONS FOR MAKING THE DISTRIBUTION OF PUPILS' SCORES, AND FOR FINDING THE MEDIAN SCORE
Interval	Number of Pupils	Interval	Number of Pupils	
160 to 169		70 & above		<p>1. The teacher must be careful that her papers are grouped correctly by classes. If she has but one grade of pupils, say 5th grade, or but two divisions of one grade, say 5th A and 5th B, then her papers are all grouped together and but one "distribution" made. If, however, she has parts of two or more grades, say part 5th and part 6th, she must make two or more piles of papers, one for each grade.</p> <p>2. Arrange the children's papers for any class group in order of the comprehension scores, the lowest score on top.</p> <p>3. To make the distribution called for, count the number of papers whose scores fall within the successive groups listed. For instance, if the lowest score is 3, the next lowest 6, the next 7, 7, 11, and so on, you will put "1" in the group marked "3 to 5"; "3" in the group marked "6 to 8"; "1" in the group marked "9 to 11," and so on until the whole number of scores are recorded. The sum of these numbers must equal the number of children taking the test.</p> <p>4. The median score is the score on the middle paper in the pile of papers arranged according to size of scores. If there are 35 papers, the median score is the score on the 18th paper. If there are 36 papers, the median score is halfway between the score on the 18th paper and the score on the 19th paper.</p> <p>5. Repeat 2, 3, and 4, for the rate scores.</p>
150 to 159		67 to 68		
140 to 149		63 to 65		
130 to 139		60 to 62		
120 to 129	1	57 to 59		
110 to 119	1	54 to 56		
100 to 109	1	51 to 53		
90 to 99	7	48 to 50		
80 to 89	5	45 to 47		
70 to 79	5	42 to 44		
60 to 69	1	39 to 41		
50 to 59	15	36 to 38		
40 to 49	2	33 to 35		
30 to 39	1	30 to 32		
20 to 29		27 to 29		
10 to 19		24 to 26	2	
0 to 9		21 to 23	1	
		18 to 20	3	
Total	39	15 to 17	1	
		12 to 14	10	
		9 to 11	9	
		6 to 8	9	
		3 to 5	3	
		0 to 2	1	
		Total	39	
		Median	10.6	

fullest extent the time and the energy of teacher and pupils are not justified. The extent to which the teacher has been trained or trains herself to complete the testing operation will determine her degree of success or failure in the use of measurements in her classroom work.

One of the first things the teacher considers after she has attained the final scores for her class is the relation which they bear to any established standards for the test which she has given.

The standards (middle of year) for the Monroe Standardized Silent Reading Tests are as follows:

GRADE	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Rate . . .	52	73	89	88	99	106	87	81	88	89
Comprehension	7.2	13	19	20	23	26.4	25	25	26.4	27.2

The scores for the 3-A class reported in Table 30 are, rate 70 and comprehension 10.6. It will be seen, therefore, that the class is above the standard in both comprehension and rate.

In the second place, the large spread of ability in rate and comprehension called for intimate knowledge with, and individual instruction to suit, these abilities.

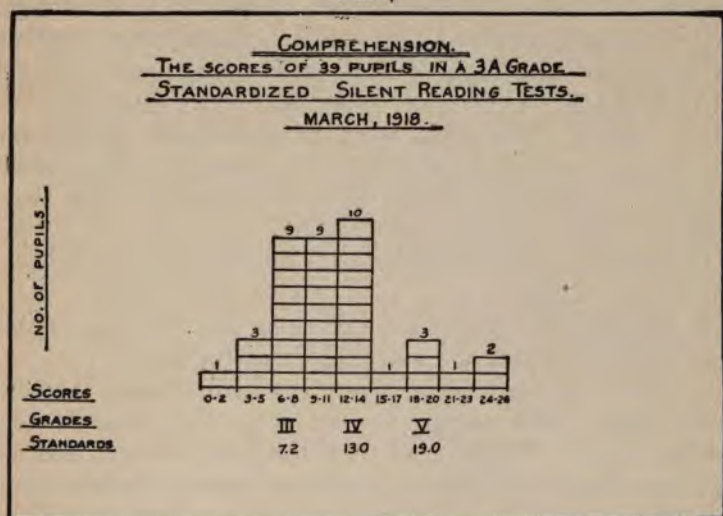
By referring to the class record sheet, it will be seen that in rate of reading at least 21 pupils surpassed the standard (52) for the third grade, 15 pupils scored above the standard (73) for the fourth grade, and 10 pupils above the standard (89) for the fifth grade. In comprehension at least 26 pupils scored above the standard (7.2) for the third grade, at least 7 pupils scored above the standard (13.0) for the fourth grade, and at least 3 pupils above the standard (19.0) for the fifth grade.

A clearer picture of this situation may be secured by



putting these results in the form of a graph as indicated in the following:

FIG. 7



Another way of seeing clearly the wide spread of ability in this class is by a comparison of the highest and lowest scores. The rate scores range from 31 words to 129 words per minute, and the comprehension scores from 1 to 26.

**Remedial Measures.** — The first problem for the teacher who gave the test reported in Table 30 was to reclassify her grade. It would be exceedingly unfair to keep the 7 pupils (using comprehension as a basis) who score above the fourth grade standard continuing in third grade reading. These 7 pupils should, therefore, either be promoted to a fourth grade (in a short time some of them may be advanced to a fifth grade) or given fourth grade reading work under the same teacher.

A second problem is emphasis on thought getting in silent reading. From the low scores in both rate and comprehension when compared with the highest scores, it is quite clear that



there are a good many pupils who do not grasp quickly the thought in what they read. The teacher who gave the test reports as follows: "In most cases those who received high rate scores also received high comprehension scores, tending to show that the rapid reader usually understands better what he reads. I, therefore, planned to give more time to silent reading. The pupils were questioned and marked on the number of questions which they were able to answer correctly." This plan helped them to understand what they read and made the lesson interesting."

In addition to the above plans, the teacher found the following devices helpful:

1. Practice was given in the selection and planning of the most interesting part of a story for dramatization.

2. Directions for reading were given by indicating thought units rather than paragraphs or lines. For example the teacher says, "William, read from the place where it tells us he (the character) decides to leave home," rather than "Read the next."

3. The pupils were given an opportunity to work up individual assignments.

4. The pupils were trained to outline the main points or events in the lesson.

5. Considerable opportunity was given for the pupils to question one another on the content without quibbling or asking silly questions.

6. Other examinations with the same tests were made to show progress.

No one can question the value of such a procedure in which the teacher gives a diagnosis of a teaching problem, and applies corrective measures in the form of intelligent and systematic practice. The plan is simple so that what this teacher has done any teacher can do. Surely the gratification of knowing without a doubt whether or not success is attending one's effort is worth while.

## SILENT READING

*Thorndike's Scale Alpha 2, for Measuring the Understanding of Sentences*

**Aim.** — The purpose of Thorndike's Scale Alpha 2 is to measure the student's ability to comprehend the meaning of sentences and paragraphs. The rate of reading is not measured.

**Description of Tests.** — The test is divided into two parts. Part I may be used in grades 3 to 5; Part II in grades 6 to 12 inclusive. The test is made up of a series of paragraphs with questions on the content of each paragraph. Each paragraph increases in difficulty over the preceding one with equal steps. Below are given the four sets of paragraphs in Part I to show the nature of the test:

## SET I. DIFFICULTY 4 (Approximately)

*Read this and then write the answers. Read it again if you need to.*

John had two brothers who were both tall. Their names were Will and Fred. John's sister, who was short, was named Mary. John liked Fred better than either of the others. All of these children except Will had red hair. He had brown hair.

1. Was John's sister tall or short?.....
2. How many brothers had John?.....
3. What was his sister's name?.....

## SET II. DIFFICULTY 5.25

*Read this and then write the answers. Read it again if you need to.*

Long after the sun had set, Tom was still waiting for Jim and Dick to come. "If they do not come before nine o'clock," he said to himself, "I will go on to Boston alone." At half past eight they came, bringing two other boys with them. Tom was very glad to see them and gave each of them one of the apples he had kept. They ate these and he ate one too. Then all went on down the road.

1. When did Jim and Dick come?.....
2. What did they do after eating the apples?.....
3. Who else came beside Jim and Dick?.....
4. How long did Tom say he would wait for them?.....

### SET III. DIFFICULTY 6

*Read this and then write the answers. Read it again if you need to.*

It may seem at first thought that every boy and girl who goes to school ought to do all the work that the teacher wishes done. But sometimes other duties prevent even the best boy or girl from doing so. If a boy's or girl's father died and he had to work afternoons and evenings to earn money to help his mother, such might be the case. A good girl might let her lessons go undone in order to help her mother by taking care of the baby.

1. What are some conditions that might make even the best boy leave school work unfinished?.....
2. What might a boy do in the evenings to help his family?....  
.....
3. How could a girl be of use to her mother?.....
4. Look at these words: idle, tribe, inch, it, ice, ivy, tide, true, tip, top, tit, tat, toe. Cross out every one of them that has an *i* and has not any *t* (*T*) in it.

*Read this and then write the answers to 5, 6, and 7. Read it again if you need to.*

Nearly fifteen thousand of the city's workers joined in the parade on September seventh, and passed before two hundred thousand cheering spectators. There were workers of both sexes in the parade, though the men far outnumbered the women.

5. What is said about the number of persons who marched in the parade?.....
6. What did the people who looked at the parade do when it passed by?.....
7. How many people saw the parade?.....

SET IV. DIFFICULTY 7

*Read this and then write the answers to 1, 2, 3, and 4. Read it again if you need to.*

You need a coal range in winter for kitchen warmth and for continuous hot-water supply, but in summer when you want a cool kitchen and less hot water, a gas range is better. The XYZ ovens are safe. In the end-ovens there is an extra set of burners for broiling.

1. What effect has the use of a gas range instead of a coal range upon the temperature of the kitchen? .....
2. For what purpose is the extra set of burners? .....
3. In what part of the stove are they situated? .....
4. During what season of the year is a gas range preferable? ...

*Read this and then write the answers to 5, 6, and 7. Read it again if you need to.*

Hay fever is a very painful, though not a dangerous, disease. It is like a very severe cold in the head, except that it lasts much longer. The nose runs; the eyes are sore; the person sneezes; he feels unable to think or work. Sometimes he has great difficulty in breathing. Hay fever is not caused by hay, but by the pollen from certain weeds and flowers. Only a small number of people get this disease, perhaps one person in fifty. Most of those who do get it can avoid it by going to live in certain places during the summer and fall. Almost every one can find some place where he does not suffer from hay fever.

5. What is the cause of hay fever? .....
6. How large a percentage of people get hay fever? .....
7. During what seasons of the year would a person have the disease described in the paragraph? .....

**Giving the Test.**—The fact that the test can be easily given to large groups of children at once makes it a useful instrument in the hands of the teacher. Before the children are asked to take the actual test, a preliminary test with instructions and subject matter very similar to the regular test is given to make sure that every child knows exactly what he is to do.

A test sheet is then placed in the hands of each child. After he has written his name and age, he is asked to follow instructions which are at the head of each paragraph. He is given as much time as is needed to do all he can.

**Recording and Determining Scores.** — The correct answers to the different questions under each paragraph have been carefully determined and arranged on an answer sheet which should be in the hands of each teacher. A portion of this answer sheet is given below:

#### KEY FOR SCALE ALPHA 2

Difficulty 4, Element 1. "Short." 2. "Two." 3. "Mary."

Difficulty 5 $\frac{1}{2}$ , Element 1. "Half past eight," "eight thirty," "8:30" or equivalent. 2. "Went down the road," or equivalent (call "went on," or "went on to Boston," wrong). 3. "Two other boys," or "two boys." 4. "Nine o'clock."

Difficulty 6, Element 1. Right responses are such as: "If he has to work afternoons and evenings to help his mother." "When their parents died." "When the father dies." "If his father died and to work." "If his father died or if sick." "His father might die." "His father may died." "If his father died he has to work." "If his father or mother died."

Etc.

This answer sheet in the hands of each teacher reduces the amount of error due to personal opinion. The rating, therefore, becomes accurate to a remarkable degree. The answer to each question is marked right or wrong according to the answer on the answer sheet.

After the marking of the papers is completed these markings are transferred to a class record sheet. This sheet has on it the number of each question under each set. It calls for the name of each child and his scores opposite under each set.

Table 31 shows a copy of a record sheet from a 5-B grade in a city school system in which the Thorndike Scale Alpha 2 was given.





If a child's answer to a question is correct the space opposite his name for this question is left vacant. If the answer is wrong a zero representing an error is recorded. The total number of errors which each child makes on each line is then determined. The value of the line on which he makes errors not to exceed 20% is his score. The total number of errors for the entire class is determined for each line. The value of the line which gives 20% of error is the score for the class. If no line gives exactly 20% of error the value of the line which gives nearest to 20% is used.

On the preceding record sheet, the class made 10.4% of errors on the line under Set II and 38.8% of errors on the line under Set III. The line which gave nearest to 20% of errors is Set II, Difficulty 5.25. By referring to a table of errors (index) it will be seen that when a line gives 10.4% of errors, .62 is to be added to the value of this line to secure the score. The score for this class is, therefore, 5.25, the value of Set II, plus .62, or 5.87.

**Interpreting the Results.** — This test is of great assistance to the teachers on account of its diagnostic value. Teaching the child to get the thought in what he reads is one of her most difficult and yet most important tasks. With the aid of this test the teacher can determine accurately his ability to comprehend the thought in a sentence or paragraph in conjunction with the class as a whole or with each member of the class.

The class record sheet given above reveals the fact that the ability of the different children to understand the thought in these paragraphs varies widely. For example M. H. and G. S.<sup>1</sup> answered each question correctly in Sets II and III. On the contrary A. S. answers three questions out of four correctly in Set II and only two out of seven in Set III. D. C. answers all the questions in Set II correctly but only one out of seven in Set III. Every teacher should keep before her the class record sheet and also each child's test sheet.

From these sheets she can tell the kind of instruction each child should receive; they will likewise tell her if he should be advanced to another grade. In fact, since thought getting is such an important factor in practically all subjects, the teacher will have considerable information about each child's general progress.

The following table shows the grade scores in a particular school in comparison with the Indiana medians:

TABLE 32

GRADES	3-B	3-A	4-B	4-A	5-B	5-A	6-B	6-A	7-B	7-A	8-B	8-A
Class scores . .	5.12		5.48	5.97	5.87	5.58	6.48	6.66	6.33	7.14	8.09	7.63
Indiana medians		5.48		6.56		7.56		8.46		8.72		9.

The above comparison reveals the fact that the reading ability in the school was below the median scores in the Indiana cities.

The general conclusion of the teachers in this building was, therefore, that the children were not able to interpret the thought in a written sentence or paragraph.

**Using the Results.** — The value of this test will be found in the direct application of the results to classroom practice. Practically every recitation affords such an opportunity. Getting the thought in what has been said or read is the big problem in all teaching. The teachers of the school in which the results in Table 32 were secured report as follows: "We stressed thought getting in all subjects such as geography, history, arithmetic, and reading. The assignments included questions calling for thought getting. The children were also asked to prepare questions calling for the thought in a certain paragraph or chapter which the members of the class were asked to answer."

One of the teachers in this school points out the following advantage of this test: "Another valuable service which these tests rendered was in determining promotions. In the June promotions the mother of a little girl who took the test questioned her child's promotion mark. When the mother was shown the rate which her child received on the reading test in comparison with the rating of other children and the answers to her questions on the different paragraphs, she was entirely reconciled to the justice of the teacher's marks. The mother also learned facts about her child which were unknown to her before. The child's answers to the questions emphasized the fact that she frequently had flights of fancy which carried her far from the point in question."

The teacher of a 4-A grade in which the test was given in March and June 1918 reported that "a class score of 5.55 was obtained in March and a class score of 6.11 in June. Some of the weaker pupils improved from 4 to 5.25 and in one case from 4 to 6." The teacher further suggests that, "If these tests were given during the first week of the term, the teacher would know in a general way how the class ranks and would also know the ability of each individual pupil in the understanding of sentences."

The Thorndike Scale Alpha 2 for Measuring the Understanding of Sentences can easily be given by the teacher. To the child it means nothing more than a class exercise in which instructions are carefully given. It measures one of the most important results of a teacher's efforts. For this reason it is strongly recommended to teachers in the teaching of reading. It is one of the best silent reading tests available.

#### COURTIS SILENT READING TEST NO. 2

**Aim.** — The aim of the Courtis Silent Reading Test No. 2 is to measure the rate and the amount of comprehension in silent reading.

**Description of Test.** — This test is published in two editions, Form I, in which appears a story entitled "The Kitten Who Played May Queen," and Form II, "The Kitten Who Went to a Picnic." Each story is of equal difficulty. The test is suitable for Grades 2 to 6 inclusive. Each test is divided into Part I and II. Part I measures the rate of reading. Part II measures the comprehension of reading.

**Giving the Test.** — Instructions to the pupils are printed on each test. These instructions are simple and easily understood by the children. Each child is permitted to read for three minutes, at the expiration of which time a line is drawn around the last word that is read. This will determine the rate of each child's reading. The child is then given five minutes in which to answer questions on what he has read. The number of questions answered correctly will determine the extent of his comprehension.

**Scoring the Results and Computing Class Scores.** — The author has provided careful instructions for the giving and scoring of the tests and the recording of the results in Folders B and D, series R, which should be in the hands of each teacher. The children in grades 5 and 6 and usually grade 4 can score one another's papers. The papers of children in grades 2, 3, and backward 4th grade classes can be scored and recorded by the children of higher grades. Answer cards are provided for this purpose.

After the median rate of reading and the median number of questions correctly answered are determined, the index of comprehension, which is "the relation the difference between the right and wrong answers bears to the right answers," is estimated. The method of determining this index of comprehension is thoroughly outlined on pages 2 to 5, Folder D. Each child's scores are then tabulated on a class record sheet. Such a record sheet for a 6-B grade in a city school system which was tested in January, 1919, is shown on pages 128 and 129.

INDEX OF COMPREHENSION  
CLASS RECORD SHEET (COURTIS)  
FRONT

City \_\_\_\_\_ School \_\_\_\_\_  
 Teacher \_\_\_\_\_ Miss H. \_\_\_\_\_  
 Room \_\_\_\_\_ Grade 6-B. \_\_\_\_\_ Date 1-29-19 . \_\_\_\_\_  
 Method \_\_\_\_\_

TABLE 1 RATE OF READING			TABLE 2 STANDARDS WORDS PER MINUTE							
Score in Words per Minute	Number of Children Making Each Score		Grade	2	3	4	5	6	7	8
Over										
400			Starch	108	126	144	168	192	216	240
380			Gray	90	138	180	204	216	228	240
360			This Test	84	113	145	168	191		
340			My City							
320			My Class					186		
300			QUESTIONS IN FIVE MINUTES							
280										
260			Grade	2	3	4	5	6	7	8
240	5		Standard	16	24	30	37	40		
220	4		My City							
200	2		My Class					28		
180	4		INDEX OF COMPREHENSION							
160	1									
140	2		Grade	2	3	4	5	6	7	8
120	4		Standard	59	78	89	93	95		
100	3		My City							
80			My Class					89		
60										
40										
20										
0										

INDEX OF COMPREHENSION  
CLASS RECORD SHEET (COURTIS)  
BACK

DIAGNOSIS		GUESSWORK			COMPREHENSION POOR ADDITIONAL TRAINING NEEDED						COMPREHENSION SATISFACTORY	
Questions Answered	Total	LESS THAN -5	-5 TO +5	6-39	40-69	70-79	80-84	85-89	90-94	95-99	100	RATE OF WORK SATISFACTORY
70												
65												
60												
55												
50	1											
45												
40	5											
35	3											
30	3											
25	9											
20	4											
15												
10												
5												
0												
Total	25				1	1	3	8	3	3	6	

Median Number of Last Question Answered 28.

Median Index of Comprehension 89.

Total Number Taking Test 25.

Number Marked I. N. F. — .



**Interpreting and Using Results.**—The preceding record is read as follows: The class is slightly below the standard for the grade in the rate of reading but considerably below the standards for the grade in number of questions answered and in the comprehension of thought. The rate of reading ranges from 100 words to 240 words in 3 minutes, a range which is entirely too great and which indicates a large number of slow readers and likewise a large number of poor readers.

From this same record sheet (back) the scores are classified according to the number of questions answered and the index of comprehension. An analysis of this classification shows only two children with satisfactory rate and comprehension and an additional four children with satisfactory rate and poor comprehension, making a total of six children out of twenty-five with a satisfactory rate. The record shows also seven children with satisfactory comprehension and poor rate, and an additional twelve children with poor rate and poor comprehension, making a total of nineteen children unsatisfactory in the rate of reading.

The standard scores for the Courtis Silent Reading Test No. 2 are as follows:

GRADE	II	III	IV	V	VI
Words per minute . . . .	84	113	145	168	191
Questions in five minutes . .	16	24	30	37	40
Index of comprehension . .	59	78	89	93	95

In order that all of the teachers in the above school could profit from the results of these tests, the following table, which compares the scores of each class with the respective grade standards, was prepared by the principal of the school and explained in detail to the teachers in conferences called for this purpose.

TABLE 33

GRADE	NO. OF CHILDREN	NO. OF WORDS READ PER MINUTE	NO. OF QUESTIONS ANSWERED IN 5 MINUTES	INDEX OF COMPREHENSION
4-B . .	61	108	25	69
4-A . .	27	149	24	84
Standard		145	30	89
5-B . .	32	137	24	86
5-B . .	26	171	28	88
5-A . .	17	122	28	94
5-A . .	12	157	32	84
Standard		168	37	93
6-B . .	25	186	28	89
6-B . .	29	169	33.5	92
6-A . .	18	154	35.5	92
Standard		191	40	95

This table is read as follows: In the 4-B grade there were 61 pupils. In this class as many pupils made a score above 108 words per minute and 25 questions answered in 5 minutes as below these scores respectively. The index of comprehension is 69. This grade is therefore below the standard in all three scores.

If the class scores are compared with class standards it will be seen that the comprehension scores with the exception of the 4-B grade (Index of comprehension 69 Standard 89) are more regular and nearer the class standards than the class scores in the other two points. The teachers of this school report: "The school as a whole measured more nearly to standard in comprehension than in the other two items. It made the poorest record in the number of questions answered in five minutes. Not a single grade reached the standard. Three fifth grade classes and one sixth grade fell below the standard for the fourth grade."

**An** analysis of the individual papers made by the teachers is **pre**sented in the following summary:

		No. of Pupils	Percentage
At <b>or</b>	above standard in all respects . . . .	7	3.6
At <b>or</b>	above standard in speed of reading . .	51	26.4
At <b>or</b>	above standard in number of questions		
answered	. . . . .	44	22.8
At <b>or</b>	above standard in index of comprehension	58	30.1
At <b>or</b>	above standard in no respect . . . .	84	43.5

The outstanding fact in this summary is the 3.6% of children scoring above and 43.5% of the children scoring below the standards for all three points.

A further analysis of the individual papers shows that "four pupils in the 5-A grade and four in the 4-B grade showed negative indices of comprehension. They wrote more incorrect than correct answers to the questions. They not only failed to get the correct meaning from what they read but obtained one absolutely opposed to the correct meaning. One of these pupils has been tested by the psychological examiner and found to be a normal child."

**Remedial Measures.**—The report from the teachers continues: "The poor showing in the number of questions answered and the almost uniformly better showing in comprehension indicates a desire for accuracy even at the expense of speed. Probably it was necessary for the children to read each paragraph several times in order to answer the questions correctly."

"It would seem that some time exercises in silent reading followed by a test for accuracy of comprehension might help to overcome this deficiency and I have recommended such exercises to the teachers to be used in the place of their regular reading lesson. Emphasis is being placed upon the time element."

In carrying out the above plan the following devices were found helpful:

1. "Expose a paragraph of reasonable difficulty on the board for a few seconds. Have the children read silently and then have them answer questions revealing their comprehension of the content of the paragraph.

2. Expose another paragraph of directions on the black-board for a few seconds and permit the children to follow them out as soon as they get the thought.

3. Expose a word picture briefly and permit the children to draw what they get from the brief exposure. Emphasize the time factor in order to show children the relative rates of reading.

4. Let a child begin to read at any point in the story. The child who discovers the place first will continue the reading."

#### ORAL READING

School practice has in the past given an undue amount of attention to the training of children in oral reading. It is not unfair to say that a much larger proportion of the school time allotted to reading is given to oral reading than to silent reading and yet, when the child becomes an adult, he will have little opportunity to practice it. Oral reading to the adult is the incidental means of expression. It is the person gifted in expression who makes a great use of oral reading. Oral reading should receive proper attention in the classroom not only as a means to successful silent reading, but also for the use that will be made of it by all individuals occasionally and a few individuals widely. It should not, however, be taught at the sacrifice of silent reading.

#### GRAY'S ORAL READING TEST

**Aim.** — The aim of Gray's Oral Reading Test is to determine accurately the extent of the child's mastery over the mechanics

of reading. This is shown by the rate of his reading and the accuracy with which he reads. The rate is determined by the number of seconds it takes to read a given paragraph. The accuracy is determined by the number of errors made in reading a paragraph. Six kinds of errors are noted, namely, complete mispronunciation, partial mispronunciation, — omissions, substitutions, insertions, and repetitions.

**Description of Test.** — The test consists of twelve paragraphs intended for grades one to eight inclusive. Each paragraph increases in difficulty over the preceding one by equal steps which have been scientifically determined.

**Giving the Test.** — Complete instructions for giving the tests are found on the back of the score sheet, which must be in the hands of each teacher using the test. These instructions should be rigidly followed. No teacher should attempt to examine her class before she has completely mastered the instructions for giving the tests and scoring the results, and until she has had some practice through the examination of two or three children. The tests can be given to only one child at a time and then not in the presence of the other children. There should be no interruptions. For this reason the test takes a much longer time for its application than is required for most tests. One teacher reports that it took her three hours and forty-five minutes to test a class of twenty-five children. This time was distributed over a number of days. The test was given after school, at noon periods and during the regular school hours in another room where there could be no interruptions. In the selection of an appropriate time for giving tests care should be taken to see that normal working conditions for the child prevail. This same teacher also reports that: "The children loved this test. I have never seen them any happier than when they were reading it for me. They liked the easy paragraphs because they were easy and they thought it was great fun to try to pronounce the difficult words in the more difficult paragraphs."



From the preceding quotation it is evident that the success with which the tests are used by a teacher depends upon the spirit with which she approaches her work and the accuracy with which she follows instructions. As the child reads from one copy of the test, the teacher follows on another copy and marks the errors as indicated on the author's instruction sheet.

**Scoring Results.** — The instructions for scoring the results are simple. The time taken to read each paragraph can be recorded in seconds on the left-hand margin of the test sheet; the number of errors made in reading each paragraph can be recorded on the right-hand margin of the test sheet. The score for each paragraph is determined from the number of seconds and the number of errors according to the following key provided by the author on the score sheet:

SECONDS	ERRORS								Or More
	0	1	2	3	4	5	6	7	
40 or more . . . . .	4	4	3	2	1	0	0	0	
30-39 . . . . .	4	4	3	2	1	1	1	0	
25-29 . . . . .	4	4	3	2	2	1	1	0	
20-24 . . . . .	4	4	3	3	2	1	1	0	
19 or less . . . . .	4	4	4	3	2	1	1	0	

"The numbers in the left-hand column refer to the number of seconds required to read a paragraph. The numbers in the horizontal line at the top of the table refer to the number of errors made in reading. The numbers in the horizontal line to the right of 40 mean that if a paragraph is read in 40 or more seconds with no errors a credit of 4 is given; with 1 error, a credit of 4; with 2 errors, a credit of 3; with 3 errors a credit of 2; etc."

The following is an actual reproduction of a child's mistakes in reading paragraph IV and the teacher's scorings:



## IV

Once there lived a king and queen in a  
 large palace. But the king and queen were not  
 happy. There were no little children in the  
 house or garden. One day they found a  
 poor little boy and girl at their door. They  
 took them into the beautiful palace and  
 made them their own. The king and queen  
 were then happy.

94  
seconds

5 errors  
(score is  
0)

It took him 94 seconds to read this paragraph and he made five errors. By referring to the key, it will be seen that the score for reading a paragraph in 40 or more seconds with five errors is zero. Therefore, his score on paragraph IV is zero. His record on all of the paragraphs is as follows:

Paragraph	I . . . . .	36"	errors	5	score	1
"	II . . . . .	48"	"	4	"	1
"	III . . . . .	86"	"	3	"	2
"	IV . . . . .	94"	"	5	"	0
"	V . . . . .	180"	"	7	"	0

A record of 86 seconds with 3 errors entitled him according to the key to a score of 2 for paragraph III. The record of 48 seconds with 4 errors entitled him to a score of 1 for paragraph II. The record of 36 seconds with 5 errors entitled him to a score of 1 for paragraph I. The teacher then entered his score for each paragraph opposite his name on the score sheet.

Below is given the exact record and also pupil score for each pupil in a class of 25 children in a 2-A grade of a city school system which was tested with Gray's Oral Reading Test.

TABLE 34.—SCORE SHEET FOR READING

ORAL READING RECORDS												
PUPIL				PARAGRAPH								
Name	Sex	Age	Nationality	1	2	3	4	5	6	7	8	Pupil Score
1. M. J. . .	F.	7	Swedish	4	4	4	4	4	4	4	2	67½
2. A. N. . .	F.	7	Swedish	4	4	4	4	4	4	2		62½
3. C. Mc. . .	F.	7	Irish	4	4	4	4	4	4	1		61¼
4. E. N. . .	F.	9	Swedish	4	4	4	4	3	4	2		61¼
5. M. W. . .	F.	8	Swedish	4	4	4	4	4	4			60
6. C. J. . .	F.	6	Norwegian	4	4	4	4	4	3	1		60
7. A. P. . .	F.	8	Swedish	4	4	4	4	3	4			58¾
8. A. R. . .	F.	7	Norwegian	4	4	4	3	4	4			58¾
9. E. H. . .	F.	7	Norwegian	4	4	4	4	4	2			57½
10. M. K. . .	F.	8	Polish	4	4	4	4	2	4			57½
11. M. Mc. .	F.	7	Irish	4	4	3	4	2	3			55
12. E. H. . .	M.	7	Swedish	4	4	4	4	2	2			55
13. E. O. . .	F.	8	Norwegian	4	4	4	4	3				53¾
14. P. M. . .	F.	8	American	4	4	4	4	3				53¾
15. R. T. . .	M.	7	Polish	4	4	3	4	1	1			51¼
16. H. H. . .	F.	7	Swedish	4	4	4	3					48¾
17. E. P. . .	F.	7	Swedish	4	4	4	1					46½
18. K. K. . .	M.	8	Polish	4	3	1	1					41¼
19. H. H. . .	M.	8	Norwegian	3	2	1	1					31½
20. L. H. . .	M.	7	Norwegian	2	2	1	2	1	1			26½
21. L. T. . .	M.	7	Polish	2	1	1	1					21¼
22. C. O. . .	F.	7	Norwegian	1	4	1	1					16½
23. J. K. . .	M.	7	Polish	1	1	2	1					13¾
24. O. R. . .	M.	8	Assyrian	1	2	1	1					13¾
25. M. M. . .	F.	10	Polish	1	1	2						12½
Total scores				83	84	76	71	48	44	10	2	
Average class score . . . . . 45.8												

In this table the initials of the children are given together with the sex, age, and nationality. It is read as follows: M. J., girl, seven years old of Swedish descent made a score

of four on paragraphs one to seven inclusive and a score of two on paragraph 8; A. N., a girl, seven years old, of Swedish descent made a score of four on paragraphs one to six inclusive and a score of two on paragraph seven, etc.

After each individual's paragraph score is determined the pupil's score for the test is found as follows: Multiply the score on paragraph I by 55 if in grade 1, 35 if in grade 2, 30 if in grade 3, 25 if in grade 4, 20 if in grade 5, 15 if in grade 6, 10 if in grade 7, 5 if in grade 8. Multiply the scores on each of the other paragraphs by 5. The sum of these products divided by four gives the score on the test.

Since M. J. was in the 2-A grade, her score for paragraph I is multiplied by 35 and each of her scores on the other paragraphs by 5. The sum of these products divided by 4 will give 67.5 according to the following process:

PARAGRAPH	SCORE	VALUE	PRODUCT
I . . . . .	4	35	140
II . . . . .	4	5	20
III . . . . .	4	5	20
IV . . . . .	4	5	20
V . . . . .	4	5	20
VI . . . . .	4	5	20
VII . . . . .	4	5	20
VIII . . . . .	2	5	10
Total product . . . . .			270

Process: Total product 270 divided by 4 equals 67.5 — score for M. J.

The average class score is found in the same manner with the exception that the sum of the individual paragraph scores is used instead of the individual paragraph scores. Referring to Table 34 it will be seen that the sum of the scores on paragraph I for this class of 25 children is 83, for paragraph II

84, for paragraph III 76, etc. Since the pupils who made scores are in the 2-A grade, the sum of their scores for paragraph I is multiplied by 35 and the sums for the scores for the other paragraphs are each multiplied by 5 according to the following process:

PARAGRAPH	SCORE	VALUE	PRODUCT
I . . . . .	83	35	2905
II . . . . .	84	5	420
III . . . . .	76	5	380
IV . . . . .	71	5	355
V . . . . .	48	5	240
VI . . . . .	44	5	220
VII . . . . .	10	5	50
VIII . . . . .	2	5	10
Total product . . . .			4580

The average class score is secured by the following process: 4580 divided by 100 (No. of children  $\times$  4) equals 45.8, average class score.

**Interpreting Results.** — The value of a test of any kind lies in the use that is made of it. Consequently, it is of the greatest importance not only that the test be given accurately but also that the results be used widely. To this end it is necessary that teachers be able to interpret their results and know how to use them.

One of the first problems for the teacher is to compare the results of her class with scores from classes in the same or other cities. Gray's Oral Reading Scale has been used in a number of surveys which have given considerable data for comparative purposes. The following table<sup>1</sup> gives the teacher an opportunity to make such comparisons:

<sup>1</sup> Courtis, *The Gary Survey*, "The Measurement of Classroom Products," p. 272.

TABLE 35

GRADE	I	II	III	IV	V	VI	VII	VIII
Gary, actual averages . . . . .		27	36	39	39	41	42	41
23 Illinois cities . . . . .		20	27	40	44	45	47	
Cleveland . . . . .		42	46	47	48	49	47	48
Grand Rapids . . . . .		44	47	49	50	48	48	48
St. Louis . . . . .		47	50	52	51	51	51	51
Gray's Standard . . . . .	31	43	46	47	48	49	47	48

This table is read as follows: On Gray's Oral Reading Test, the second grade pupils in Gary made an average score of 27; in 23 Illinois cities, 20; in Cleveland, 42; etc.

Since the average score of the second grade class reported in Table 34 was 45.8, this class scored higher than Gray's Standard and every city score except St. Louis. Such comparisons give the teacher information which enables her to base her practice on scientific facts rather than on opinion.

The teacher of this class says in this connection: "I expected it to come out that way because I think this class as a whole is doing good work in oral reading. Some of the children are very unusual readers, and there are not so many poor readers." The test in this class also reveals the fact that there are a few children who are very poor oral readers and the extent to which they are below the average for the class. It, therefore, becomes a means of dividing a class of students with reference to their ability. It is here that the test reveals its greatest value. While it is important to know just where a class stands with reference to average ability in a certain subject, it is far more important to know the attainment of each student in that particular subject. In this way practice can be so regulated that it meets the needs of each individual and does not result in failure to both teachers and students. It often happens that the teacher does not form a correct judgment of a child's ability. This is illustrated in

the case of E. N. (Table 34), about whom the teacher makes the following report: "E—— made a high score in this test and I think the test was valuable for that reason in that it showed me how much E—— really can do. The children don't look pleased in class when it is E——'s turn to read because she reads in such a monotonous way although I have worked very hard with her. One would never give E—— credit for being one of the best oral readers, but she has proved by this test that she does know the words and the mechanics of reading."

Again the tests will determine accurately the best readers in the class. Concerning the five students (Table 34), who made the best scores the teacher reports: "These are my best readers. The test proved this very accurately."

The report of the teacher also reveals the fact that too much care cannot be exercised in seeing that normal conditions surround the child when the test is taken. If the child is interrupted or if it is made to feel that undue importance is attached to the result, nervousness may greatly hinder a true statement of the child's ability. In the case of C. O. (Table 34), who made a score of 16.25, the teacher reports: "C—— made a poor score; . . . we all love to hear C—— read and I consider her a good oral reader. I think she seemed a little nervous for fear she wouldn't do as well as the others and she made so many little mistakes, which brought her score down and which she seldom does in school." The teacher also makes the same explanation for the low score of L. H. She says: "L—— made a poor score but he is one of the best oral readers in my room. He was so anxious to excel, and I think that made him nervous, for one would expect him to stand at the head instead of at the foot." This shows the need to have the tests given under normal conditions.

**Using the Results.** — Permanent progress resulting from the use of tests will depend upon the use that is made of them. Consequently, careful attention should be given to the following work:



First, the test should be given at the beginning and at the end of the term so that time and energy of pupils and teachers are not wasted in finding out what children can and cannot do.

Second, a graph of each individual score should be kept in a convenient place so that each child can see his standing in relation to his classmates. The following is a convenient graph to use in connection with class results:

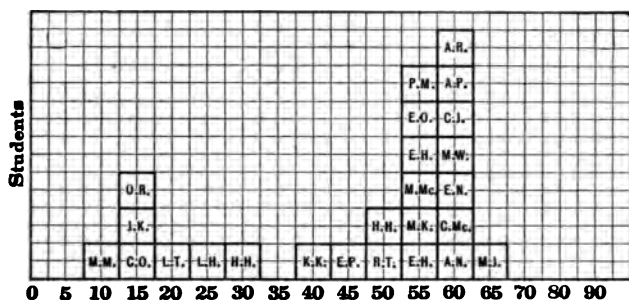


FIG. 8.— Showing the position of each student in the 2-A grade (Table 34) according to his oral reading ability.

Third, the teacher should keep each child's test sheet in order that his difficulties in the mastery of symbols in reading may be investigated.

Fourth, the children in the class should be grouped into fast and slow groups according to their ability as revealed by the test.

Fifth, the wide variability in the achievement of children in practically every class calls for serious consideration by the teacher in the way of readjustment of class groups, special promotions, etc.

If many of these children (Table 34) are good in the other subjects of the grade they should be given an opportunity to advance to the work of the third grade. The question should be asked: Are not some of these children being held back for the slower children? Numerous cases are on record to show that when such children are given an opportunity to advance

to a higher grade, they are able to maintain the standard of the grade without much difficulty and advance with the class, much to the surprise of the teacher.

### HAGGERTY'S VISUAL VOCABULARY TEST

**Aim.** — This test aims to determine the extent to which the children have acquired control of words. It is specially helpful to the teacher in the early stages of reading.

**Description of the Test.** — The Haggerty Visual Vocabulary Test is arranged in two parts. The second part is treated later in the chapter. The first part is a test for grades one and two and is divided into series A and B. Each series consists of 30 sight words and 25 phonetic words arranged in lines. Each line bears a number. The words in series B are different from those in series A but are of the same degree of difficulty in order that a class may be tested a second time to determine the amount of progress during a certain period of time. These tests are printed in the following form :

#### PUPIL'S CARD

<i>Sight</i>		<i>Phonetic</i>	
50	come	50	bit
	one		cow
	who		that
	she		out
	on		fox
75	mamma	75	yet
	next		trick
	blue		toy
	wood		frog
	rabbit		find
55	your	55	stay
	pretty		ran
	yes		gun
	too		like
	house		meat
85	lion	85	crown
	monkey		chew
	cradle		kite
	naughty		snag
	visit		lace
65	here	65	hand
	has		ten
	bird		name
	put		head
	shall		cold
95	hurrah		
	pigeon		
	circus		
	picnic		
	pieces		

**Giving the Test.** — The application of the test is simple so that it can easily become a helpful instrument in the hands of the teacher. The following instructions are sufficient to insure good results :

1. Each pupil is to be tested alone.
2. Hand the child the Pupil's Card.
3. Ask him to pronounce each word beginning at the top of the column.
4. Do not help the child in any way. Do not correct mistakes or suggest ways of working out a word. Do not suggest that the child has seen the word before. Do not seem impatient if the child makes an error. Allow a reasonable time for each word, and if the child does not name the word correctly, ask him to try the next word.
5. Whenever the pupil fails to speak the word correctly place on the Class Record Card a zero opposite the word and in the column allotted to that pupil.
6. Record the age of each pupil to the nearest month, that is, 6: 5 for six years five months.
7. Record boys by letter B; girls by letter G.

**Scoring Results.** — When all the pupils of the class have been tested, total the number of zeros or errors on each line for the class and find the percentage this amount is of the total number of scores for each line. If fifteen children are taking the test there should be 75 scorings ( $15 \times 5$ ) on a line of five words. If 5 of these scorings are errors, the percentage of error would be  $5 \div 75$  or  $6\frac{2}{3}\%$ . In the same way each child's percentage of errors is determined.

"The highest numbered line which the child does with one (or no) omission or error is taken as his score."

The score for the class will be the number of the line which has 20% of errors. "If no single line gives exactly 20%, the actual class score will be intermediate between the two lines which give nearest 20% of error." (See Form 9, series B, Author's Directions for Giving and Scoring Tests in Reading.)

**Interpreting the Results.** — The teacher giving the Haggerty Visual Vocabulary Test should have no difficulty in interpreting the results of her test and in applying them to her practice in the classroom. The results are usually pronounced, due to the fact that the progress of children in the mastery of words is rapid if proper methods are employed. The results of the work of five teachers in the first and second grades of a city school system are reported in Table 36, which is read as follows: Thirty-five children in the 1-A grade of the Monroe School made a score of 37.3 in February and 44.9 in May on sight words, which is a growth of 7.6 points; twenty-two children in the 2-B grade of the same school made a score of 50.7 in February and 76.2 in May, which is a growth of 25.5 points, etc.

TABLE 36. — THE RESULTS OF THE HAGGERTY VISUAL VOCABULARY TEST GIVEN IN GRADES I AND II IN TWO SCHOOLS IN FEBRUARY AND AGAIN IN MAY, 1918

SIGHT WORDS									
Monroe					Bryant				
GRADE	ENROLL- MENT	SCORE FEB.	MAY	DIF.	GRADE	ENROLL- MENT	SCORE FEB.	MAY	DIF.
1-A .	35	37.3	44.9	7.6	1-A .	21	42.4	52.4	10.0
					1-A .	34	29.3	39.1	9.8
2-B .	22	50.7	76.2	25.5	2-B .	32	56.0	67.6	11.6
2-A .	15	69.0	89.7	20.7	2-A .	31	82.7	92.7	10.0
2-A .	13	75.8	73.9	-1.9	2-A .	10	81.3	84.8	3.5

PHONETIC WORDS									
GRADE	ENROLL- MENT	SCORE FEB.	MAY	DIF.	GRADE	ENROLL- MENT	SCORE FEB.	MAY	DIF.
1-A .	35	29.3	50.2	20.9	1-A .	21	39.2	58.5	19.3
					1-A .	34	30.2	38.4	8.2
2-B .	22	46.9	79.	32.1	2-B .	32	46.6	70.7	24.1
2-A .	15	66.1	87.4	21.3	2-A .	31	77.	87.9	10.9
2-A .	13	68.9	71.7	2.8	2-A .	10	70.	77.	7.

The column of differences shows a wide variation in the amount of growth, which is due to the teaching, the mental condition of the class, changes, etc. The principal under whose supervision these tests were given reports: "The least progress was made in the grades in each school where the teachers were changed during the semester. Their scores furnish a mathematical statement of the loss caused by such a change."

If the results of these tests as reported in Table 36 are represented in graph form, a better idea of the amount of growth in each class can be more readily secured.

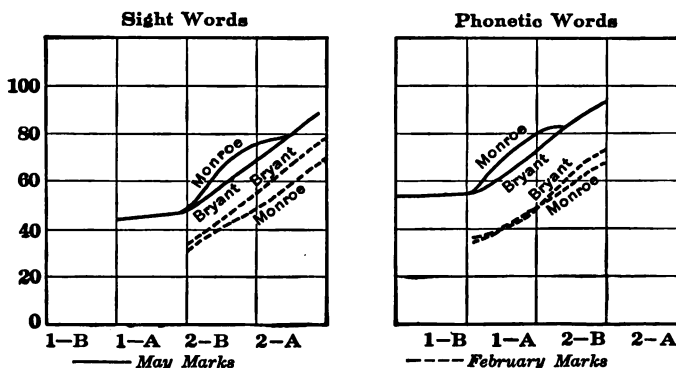


FIG. 9. — The amount of growth in the mastery of sight and phonetic words in grades I and II of two schools from February to May, 1918.

The graphs show that the classes have made approximately a half year's progress. The Monroe School occupies a slightly better position in May than it did in February.

**Using the Results.** — The use which can be made of these tests and the extent to which they can affect classroom practice is best told in the words of one of the teachers who gave the test. "The fact that so many children failed to recognize such words as name, head, think, here, yes, shows that more drill should be given to words of that type." "An interesting



feature of the test was the manner in which different pupils gave the words. Those who were sure of the words gave them rapidly. Many knew the phonograms, but failed to enunciate them properly. Some who read well in their books failed to recognize the words because they were isolated."

"After studying the results I decided to adopt a more efficient method of teaching phonics.

"1. I reviewed carefully the basic facts of phonetics and made plans for a more complete course of phonetic instruction.

"2. I determined to give more attention to individual instruction in the mechanics of reading. The number of concert recitations was limited, and the classes were grouped according to weaknesses along important lines. I found that some children needed more drill on the basic phonograms, and others on proper enunciation.

"3. At the suggestion of the primary supervisor a system of checking up daily the individual errors was tried out by means of the following chart:

PHONIC CHART											
Pupil's Name	Melba	Lorena	Wilbert	Ames	Alvin	Hildur	Ester	Tony	Viola	Carl	Grace
Phonograms	100		100	100	100		100		100	100	100
AY											
AKE		✓									
AME						✓					
ADE								✓			
AIN								✓			
AVE											

"As each child pronounced the phonograms on the left side of the chart his record was marked in the column with his name at the top. If no errors were made, one hundred was placed under his name; if errors were made a check was placed opposite the phonogram.



"In conclusion, the tests emphasize the importance of individual help in teaching phonics. It has brought to me very strongly the fact that individual help in all other subjects as well is equally important."

**Corrective Measures.** — Many reading recitations would be more profitable if they were preceded by a study period in which the children followed a definite assignment. Too often a class is hastily told to read seven pages or the next story and be able to tell what has been read. This is wholly inadequate, leading to a hurried and aimless sketching in which the child unconsciously omits or neglects the very things he should master and spends his time on the easy and entertaining parts because he has not been taught to analyze the difficulties.

A definite assignment is necessary to insure a profitable study period. The nature of the subject matter and the status of the class determine the kind of assignment that will be most effective.

The following methods have been found helpful:

1. A word and phrase drill should precede the study period with emphasis on the *time factor*, — quick recognition is the important thing. Pupils may practice during the study period to reduce the time it requires them to recognize these lists. Group work may be done.

2. Questions and directions on the board will lead to a study of thoughtful interpretations. For example, "In which speech is John angry?" "Is he slightly vexed or very angry?" "Show by your reading which you think he is."

3. A short enunciation drill may be assigned for study before class.

4. Individual assignments to correct individual difficulties may be given.

## OTHER TESTS

*The Kansas Silent Reading Tests* devised by Dr. F. J. Kelley<sup>1</sup> are made up of a graded list of paragraphs or exercises. These tests preceded and are not unlike the Monroe Standardized Silent Reading Tests. The Monroe Tests have incorporated those features of the Kansas Silent Reading Tests which have proved satisfactory. There are three tests: No. 1, for grades 3, 4, and 5; No. 2, for grades 6, 7, and 8; No. 3, for grades 9, 10, 11, and 12. Each test is divided into two forms, which are different in subject matter, but equal in difficulty, so that the same grades can be measured at different times. The tests are intended to measure speed and comprehension in silent reading. The scores on these two points are given in one mark. The score in speed and comprehension for each exercise is recorded on the left-hand margin of the test sheet. The child's ability in silent reading is, therefore, the sum of the values of the exercises which he reads correctly in five minutes.

These tests have been widely used so that standards are available. These standards are given in terms of the median and percentile scores as indicated below:

GRADE	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Twenty-five percentile	2.5	6.1	9.4	9.4	11.8	13.7	16.0	17.9	18.7	22.3
Median score . . .	5.3	9.5	13.0	13.0	16.2	19.2	22.9	25.6	26.5	29.7
Seventy-five percentile	8.2	13.6	17.5	19.8	21.9	26.4	30.4	31.9	33.1	34.1

*The Fordyce Scale for Measuring the Achievements in Reading*, devised by Dr. Charles Fordyce,<sup>1</sup> is intended to measure the speed and quality (comprehension) of silent reading. The scale is divided into test No. 1, designed for grades 3, 4, and 5, and test No. 2, designed for grades 6, 7,

<sup>1</sup> See Bibliography.

8, and 9. The legend of "Narcissus" is the selection for test I and "The Spirit of Spring" the selection for test II. The rate of reading is determined by the number of words read at the expiration of three minutes for Part I and five minutes for Part II. The extent of the pupils' ability to understand is determined by answering certain questions on the entire selection for which ten minutes and fifteen minutes are given respectively for Tests I and II. In order that all the children may have the same information, they are given an opportunity to finish the part which they have not yet read. Below are given the standards in percentages for the test:

Test No. I, designed for Grades III, IV, and V.

GRADE	III	IV	V
Speed . . . . .	90	95	100
Quality . . . . .	57	71	74

Test No. II, designed for Grades VI, VII, and VIII

GRADE	VI	VII	VIII
Speed . . . . .	90	100	100
Quality . . . . .	73	75	76

This test is arranged in convenient form so that the teacher can give it to her class quickly and efficiently.

It gives the reading ability of children in terms of speed and quality (comprehension). The score sheet is arranged so that the ability of each child can easily be seen. The scale becomes, therefore, an important instrument in the hands of the teacher to diagnose the needs of her class. It is receiving wide use throughout certain sections of the country.

*Brown's Silent Reading Tests*, constructed by H. A. Brown,<sup>1</sup> are made up of easy reading selections to be used in grades 3 to 8. They are intended to measure speed and the quantity and quality of comprehension. After the children have read as much as they can in exactly one minute they are asked to write as much as they can remember of what they have read. A key is used to determine the quantity and the quality of what has been reproduced. In addition to the scores in speed and comprehension, the child's ability is given in terms of one mark called "reading efficiency," which is obtained by multiplying the scores in speed and comprehension. The following are tentative scores so far available for the test:

	WORDS PER SECOND	COMPREHENSION	READING EFFICIENCY
Grade III . . . . .	3.32	46	127.8
Grade IV . . . . .	3.55	65	217.1
Grade V . . . . .	4.40	61	291.0
Grade VI . . . . .	4.54	68	295.0
Grade VII . . . . .	4.65	78	322.3
Grade VIII . . . . .	4.84	79	323.6

The tests are useful in determining the reading ability of individual children on account of the fact that the different factors — speed, quantity, and quality of comprehension — which are necessary for good reading can be determined.

*The Haggerty Visual Vocabulary Tests* by Dr. M. E. Haggerty are a "slight modification" of the Thorndike Visual Vocabulary Tests with the addition of an oral test (Part I) for grades I and II which is discussed under Oral Reading.

Part II is a list of words called "Scale R" for grades III and IV and another list of words called "Scale R 2" for grades V, VI, VII, and VIII.

<sup>1</sup> See Bibliography.

*The Thorndike Reading Scale, Word Knowledge or Visual Vocabulary*, by Dr. E. L. Thorndike is intended to determine the extent of a child's knowledge of words.

The scale is divided into two divisions, Scale A-2 and Scale B. Each scale has two series, X and Y. Each series is made up of a graded list of words which increases from simple words familiar to almost any child with two or three years in school to less familiar words which school children seldom meet. The list of words for each series and for each scale is of equal degree of difficulty for the purpose of testing the same children more than once to determine the extent of progress. These tests are intended for grades III to VIII.

*Gray's Silent Reading Test* by Dr. Wm. S. Gray<sup>1</sup> is made up of the following selections for grades as indicated:

Grades II and III	Tiny Tad
Grades IV, V and VI	The Grasshoppers
Grades VII and VIII	Ancient Ships

The test is used to determine the rate and quality of pupils' reading. Quality is "based on the ability of pupils to reproduce what was read and to answer specific questions concerning the subject matter of the test." Only one pupil is tested at a time. This test has recently received considerable use in the St. Louis survey. On account of the large amount of time required to give this test, its use is limited in the hands of classroom teachers. The standard scores for this test are as follows:

GRADE	II	III	IV	V	VI	VII	VIII
Rate (words per second)	1.50	2.30	2.20	2.57	2.79	2.69	2.87
Quality . . . . .	32	37	29	32	39	22	27

<sup>1</sup> See Bibliography.



*Achievement Examination in Reading: Sigma 1* by Dr. M. E. Haggerty and Margaret E. Noonan is intended to measure the reading ability of children in grades I to III inclusive. It was devised in 1919 in connection with the Virginia school survey. In addition to its use in the state of Virginia, it has been used in St. Louis, Cincinnati, Madison, Bloomington, and Aberdeen.

This reading test is in two divisions in one pamphlet, Test 1 and Test 2. Test 2, which consists of a series of questions to be answered by no or yes, as "Can you eat? No—yes; Can a bat walk? No—Yes," etc., should be given first. Test 1 consists of a series of performances, as "Put a tail on this pig." (Picture of pig without tail given.) A Manual of Directions by Dr. M. E. Haggerty contains explicit instructions for this test and also for Intelligence Examinations: Delta 1 and Delta 2. This manual should be in the hands of every teacher giving the test.

The results from this reading test may be secured in terms of scores for grades and age groups. The following grade standards and age norms are available:

TABLE 37. — GRADE STANDARDS FOR ACHIEVEMENT TEST IN READING.  
SIGMA 1

GRADE		I	II	III	IV
Score {	Test 1 . . . . .	4	12	16	20
	Test 2 . . . . .	2	8	14	18

TABLE 38. — AGE NORMS FOR ACHIEVEMENT TEST IN READING.  
SIGMA 1

AGE IN YEARS		7	8	9	10	11
Score {	Test 1 . . .	6	12	15	18	24
	Test 2 . . .	4	7	12	15	19



The authors of these tests have overcome the difficulties encountered in determining the reading ability of young children who are unable to follow complicated instructions or who have not acquired a wide vocabulary. The nature of the subject matter appeals to the interest of the children. The simplicity of the tests and the ease with which they can be given make them a valuable instrument in the hands of every primary teacher who wishes to know the achievement of her children in reading.

The subject of reading has been receiving marked attention from psychologists and other educational experts during recent years. The importance of reading in every child's development justifies this attention. This interest has resulted in sufficient scientific information in the form of standardized tests, scales, and standards of achievement which enable every classroom teacher to determine accurately the ability of her children in reading. Every teacher now in a classroom or coming fresh from the training school should be sufficiently familiar with a test or several tests so that she can justify her instruction in reading by scientific facts as well as by opinion.

#### BIBLIOGRAPHY

- Fordyce, C. F., "Scale for Measuring the Achievements in Reading." Price, Test 1 or 2 complete with record sheets and directions, \$1.25 per hundred pupils if the practice exercises are included; \$.75 for a hundred pupils without the practice exercises. Address the University Publishing Co., Chicago, Illinois.
- Gray, William S., "Gray's Silent Test." Oral-reading tests with directions and score sheets, \$.50 per hundred. Silent Reading tests, \$.50 per hundred. Postage or express extra. Address, W. S. Gray, School of Education, University of Chicago, Chicago, Illinois.
- Haggerty, M. E., "Visual Vocabulary Test. Visual Vocabulary for Grades I and II." Phonetic or sight cards (either), \$.00½ each; record sheets, \$.01½ each; score sheets, \$.03 each; directions, \$.01½ each. Address, Northwestern School Supply Co., Minneapolis, Minnesota.

- Haggerty, M. E., and Noonan, Margaret E., "Achievement Examination in Reading: Sigma 1." World Book Co., Yonkers-on-Hudson, New York City. Manual of Directions, \$.35. Test, \$1.40 per package of 25. Scoring Key, \$.05.
- Kelley, F. J. "The Kansas Silent Reading Tests." Price, including directions and record sheets, \$.50 per hundred. Address, Bureau of Educational Measurements and Standards, Kansas State Normal School, Emporia, Kansas.
- Monroe, W. S., "Monroe's Standardized Silent Reading Test." Price, including directions and record sheets, \$.60 per hundred. Address, Walter S. Monroe, Indiana University, Bloomington, Indiana.
- Thorndike, E. L., "Improved Scales for Word Knowledge or Visual Vocabulary." Scale A 2 and Scale B. Printed in four sheets: Scale A 2,  $x$  series; Scale A 2,  $y$  series; Scale B,  $x$  series; and Scale B,  $y$  series. Price each \$.40 per hundred; \$3.25 per thousand of any one kind. Postage extra. Sample set, \$.06 by mail. "Improved Scale for Measuring the Understanding of Sentences," Scale Alpha 2, parts 1 and 2. Price, each, \$.50 per hundred. Record sheets, \$.25 a dozen. Postage extra. Sample set, \$.08 by mail. Address, Bureau of Publications, Teachers College, Columbia University, New York City.

## CHAPTER VI

### THE MEASUREMENT OF ENGLISH COMPOSITION

THE importance of written composition in the life of the pupil and the indefinite notions of teachers as to proper standards in written composition make the need for an objective measurement an exceedingly urgent one. The difficulty of securing such a measurement in written composition over the measurement of such subjects as spelling, arithmetic, penmanship, is greatly increased by the large number of factors which make up written composition. There are the different kinds of written composition, as narration, description, exposition, and argumentation; there are also the factors of the content and such form elements as punctuation, spelling, capitalization, and sentence structure. All of these difficulties have hindered the construction of a scale for written composition which can be used with ease and accuracy by the classroom teacher in determining the ability of her children in this subject.

The first scale for written composition was constructed by Dr. Milo B. Hillegas.<sup>1</sup> It measures general merit in written composition. It consists of 10 samples, 3 of which are artificial, 5 were written by high school students, and 2 by college freshmen. The values of the different samples range from 0 to 9.3 with wide steps between samples. This scale has shown the need for a written composition scale and formed the basis for the construction of other scales which have a more direct application to particular grades of work or to local

<sup>1</sup> See Bibliography.

situations. Examples of this procedure are found in the Nassau County Supplement to the Hillegas Scale for Measuring the Quality of English Composition, proposed by Dr. M. R. Trabue<sup>1</sup> and the Extension of the Hillegas Scale for Measuring the Quality of English Composition of Young People by Dr. E. L. Thorndike.<sup>1</sup>

THE NASSAU COUNTY SUPPLEMENT TO THE HILLEGAS  
SCALE

**Aim.** — The purpose of this scale is to determine the general merit of children's written composition. No attempt has been made by the author to define the different elements in general merit. Hillegas in referring to his own scale which measures the same thing says, "The term (merit) means just that quality which competent persons commonly consider as merit, and the scale measures just this quality."

**Description of Tests.** — For the purpose of giving the teacher a clearer understanding of the scale and its application, the entire scale is given below:

WHAT I SHOULD LIKE TO DO NEXT SATURDAY

O

I went going on to the Dox Saturdaye dnd day we the boys and I well going home and I well going the boys. and I will going these read in and they to night. and we or night. I well going a ground shalt and I gone out I will going to shea shouse and I will shoe or the skill of the shea of night.

I.I

I intend to mak a snou man and make an fort and fort snou ball at chidern and hau I whist ma frant carolyn cole what were me I will going to the mauiss on Saturday. Georga will come went me.

<sup>1</sup> See Bibliography.

at night I will going out went my mother to the marce. I will mak the snou man and the fort in the moning and in the afternoon I will go to the mauies. I whist there whest school on Saturday.

## 1.9

one next Saturday. I expect to go to the city leve next Gaturday to see my ofriend archie king I am going to grow to the baning balys circus with hime next Saturday before I go I have to do my jobs feedsing the cows and horse ard chinkens and geese next Saturday. My friend is a very good fellow to go and see So my mother Said "If I do my work during Easter week vacation I can go to the barning baley circus with. hime

## 2.8

Once a pon a time there was a girl. One day she asked me what I was going to do next Saturday so I said, "I am going to go for a swim." And she said, "thats just where I am going to." next Saterdag came we both went down together. We came home at noon time. after dinner we went to the picktures. There we had a good time. And then came home at night.

## 3.8

I would like to go out in the afternoon and play catching the ball. Go over to Bertha's house and have a few girls to come with me and be on each others side. I have a tennis ball too play with. The game is that one person should stand quite aways from another person and throw the ball too one then another. Someone has to be in the middle and try too get the ball a way from someone then she takes this persons place who she caught the ball from. Then till every person has a chance.

## 5.0

Next Saturday I should like to go away and have a good time on a farm. I should like to watch the men plowing

the fields and planting corn, wheat, and oats and other things planted on farms.

Next Saturday I will go to the Pioneer meeting if nothing happens so that I cannot go. I should like to go swimming but it is not warm enough and I would catch a bad cold. I should like to go to my aunts and drive the horses, I do not drive without some older person with me, so I cannot go very often.

I should like to see my aunts cat and her kittens, too. I think I can, to.

### 6.0

I should like to join my girl friends, who are going to the city on the 9-05 A.M. train. They are going shopping in the morning and will have lunch to-gether, then they are going to the Hippodrome. After the Hippodrome, they are all going home to dinner to one of the girls houses, she lives on Riverside Drive so they expect to take the "Fifth Avenue Bus" up there. The evening will be devoted to playing games, singing and dancing.

### 7.2

If I had a thousand dollars to spend, I think I would take a trip to San Francisco by train with the rest of the family, and stop at a sea-side hotel. It would be glorious to see the surf again, and to escape from the cold blustering weather of December for the balmy breezes of the ocean, and the whiff of orange blossoms.

We could take long drives under shady trees, visit the orange and olive groves and bathe in the surf. Think of bathing in the ocean in December.

Coming home again I would enjoy stopping at Yellow Stone Park. It would be lots of fun to camp out, and to ride over the prairies on frisky ponies. It would be very interesting to notice the change of climate as we got farther east, and to go to bed on the train one evening feeling warm, and waking up the next morning feeling very chilly.



I am afraid by the time I would get home a thousand dollars would be pretty well used up; but if not I would like to give a party.

### 8.0

One Sunday, towards the end of my summer vacation, I was in bathing at the Parkway Baths. In the Brighton Beach Motor drome, a few rods away, an aviation meet was going on. Several times one of the droning machines had gone whirring by over our heads, so that when the buzzing exhaust of a flier was heard it did not cause very much comment. Soon, however, the white planes of "Tom" Sopwith's Wright machine were seen glimmering above the grandstand. Everyone stood spellbound as he circled the tract several times and then headed out to sea. He was seen to have a passenger with him. Suddenly, the regular hum of his motor was broken by severe pops, and the engine ran slower, missing fire badly. In response, to Sopwith's movements, the big flier tilted and swooped down to the beach from aloft like an eagle. The terrified crowd make a rush to get out of the way as the airship came on, but Sopwith could not land on the beach, but skimmed along close to the water instead. Suddenly his wing caught the water, and the big machine somersaulted and sank beneath the waves. The aviators soon came bobbing up and were taken away in a launch, but the accident will not soon be forgotten by those who saw it.

### 9.0

The courage of the panting fugitive was not gone; she was game to the tip of her high-bred ears; but the fearful pace at which she had just been going told on her. Her legs trembled, and her heart beat like a trip-hammer. She slowed her speed perforce, but still fled industriously up the right bank of the stream. When she had gone a couple of miles and the dogs were evidently gaining again, she crossed the broad, deep brook, climbed the steep left bank, and fled on in the direction of the Mt. Marcy trail. The fording of

the river threw the hounds off for a time; she knew by their uncertain yelping, up and down the opposite bank, that she had a little respite; she used it, however, to push on until the baying was faint in her ears, and then she dropped exhausted upon the ground.

The first seven specimens were selected from compositions written by children in the elementary schools of Nassau County, New York, on the topic, "What I Should Like to Do Next Saturday." The last three were selected from a list of compositions published by Dr. E. L. Thorndike. The values range from 0 to 9. It is intended for grades 4 to 12 inclusive. The compositions are arranged on one sheet with the value of each composition printed on the left-hand margin. For the sake of clearness in this text the value of each paragraph is placed in the middle of the page and above the paragraph.

**Applying the Scale.** — After a set of composition papers has been secured which has been a part of the regular class work or a special assignment on some topic which is familiar to the pupils as an Exciting Experience, An Accident, What I shall Do Next Saturday, etc., each paper is compared with the scale with a view of deciding on the specimen which its quality most closely resembles. The scale value of this composition is marked on the child's composition. If a finer rating is desired than the values as given to each composition, units between each value may be used. For example if a composition is better than quality 3.8 and not as good as quality 5.0 a rating between these two qualities may be given as 4 or 4.5 according to the judgment of the teacher.

**Scoring the Results.** — As soon as a rating has been given to each child's composition the papers are grouped according to their values, from the lowest to the highest. The number of papers in each group is then determined and the results recorded on a record sheet similar to the following:

TABLE 39.—RESULTS IN A SCHOOL GRADES IV TO VI

## CLASS RECORD SHEET

City D.

School L.

Grades 4-B—6-A

Teacher or Principal

Date Jan. 7, 1920

RATINGS	NUMBER OF SCORES IN GRADES																	
	IV		V		VI		VII		VIII		IX		X		XI		XII	
	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A
0. . . .	7	5																
1.1 . . .	6	6	1															
1.9 . . .	7	7	8	2	1													
2.8 . . .	2	2	19	16	1	1												
3.8 . . .	1	1	11	14	10	9												
5.0 . . .			3	11	18	9												
6.0 . . .					8	14												
7.2 . . .																		
8.0 . . .																		
9.0 . . .																		
Total No. of papers	23	21	42	43	38	33												
Median class scores .	1.1	1.1	2.8	3.8	5.0	5.0												

After the rates for all the compositions are recorded on the class record sheet, the median class score is determined.

The above table is read as follows: In 4-B grade there are 23 pupils of which 7 received a rating of 0, 6 received a rating 1.1, etc. The median score for the 4-B grade is 1.1.

**Interpreting and Using the Results.**—The tentative standards for the Nassau County Supplement as arranged by the author are as follows:

TABLE 40

GRADE	TENTATIVE MEDIAN STANDARD
Fourth . . . . .	3.5
Fifth . . . . .	4.0
Sixth . . . . .	4.5
Seventh . . . . .	5.0
Eighth . . . . .	5.5
Ninth . . . . .	6.0
Tenth . . . . .	6.5
Eleventh . . . . .	6.9
Twelfth . . . . .	7.2

The following table taken from the Nassau County Survey shows the scores which have been attained in the cities named :

TABLE 41

SCHOOL SYSTEM	MEDIAN SCORE ATTAINED IN GRADE									
	IV	V	VI	VII	VIII	IX	X	XI	XII	
Nassau County . . . . .	2.76	3.42	3.82	4.18	4.56	5.00	5.25	5.68	5.94	
Lead, S. D. . . . .	3.57	4.11	4.64	5.01	5.57					
Newark, N. J. (one school) .	2.39	2.51	3.56	4.33	5.27					
Ethical Culture School, N.Y.		4.01	4.72	5.39	5.74					
Chatham, N. J. . . . .	2.95	2.85	4.10	4.02	5.29					
Salt Lake City . . . . .	3.58	3.84	4.61	5.16	6.37					
Butte, Mont. . . . .	2.34	2.80	3.41	3.77	4.11					
South River, N. J. . . . .	2.31	2.33	3.78	4.75	5.62	5.18	5.02	5.95	6.30	
Mobile County, Ala. . . . .	3.20	3.91	4.34	4.22		5.56	6.38	6.05	6.77	
Mobile, Ala. . . . .	3.31	3.85	4.60	4.95		6.69	6.93	7.24	7.54	
54 high schools . . . . .						4.99	5.88	6.38	6.69	

The following table shows the median scores for rating compositions of 539 pupils in grades 4 to 6 inclusive for four schools in a city school system in January, 1920.

TABLE 42

SCHOOL	IV-B	IV-A	V-B	V-A	VI-B	VI-A
1 . . . .			2.8	3.8	3.8	5.0
2 . . . .			2.8	2.8	2.8	3.8
3 . . . .	1.1	1.1	2.8	3.8	5.0	5.0
4 . . . .	2.8	2.8	3.8	3.8	3.8	3.8

If a comparison is made of the scores reported on the class record sheet (Table 39) with the scores in Table 41 it will be seen that the written composition work in this school in the fourth grade is below all the scores, in the 5-B grade it is below all scores except two, in the 5-A grade it is exceeded by 5 out of 10 scores, and in the 6-B and 6-A grades it exceeds all scores.

It is evident, therefore, that the written composition in the fourth grade is exceedingly low, in the fifth grade it is also low, but in the sixth grade it is ahead of the work in the places with which comparisons are possible.

The unusually low scores in the fourth grade may be explained partially by the large number of foreign children in this school. These scores are surely an indication that more systematic training in oral and written composition should be given in the fourth grade and also in the third grade.

The scores in Table 42 were obtained from compositions which were written as a class exercise under the supervision of the classroom teacher. All of the children wrote on the same theme, "An Exciting Experience." Twenty-five minutes were allowed in which to do the work. The exercise was given one week before the end of the first semester. The purpose of the exercise was two-fold: first, to determine the attitude of the teacher towards the use of such a scale in rating composition papers as opposed to the regular method; and second, to ascertain whether or not the scale in con-

nection with such an exercise could be used to determine promotions.

The teachers who gave the exercise and scored the papers were unfamiliar with the use of the scale. Carefully prepared instructions were given to each teacher.

The results show that in general the composition work in these four schools in comparison with the standards is low. Judging from the comments of the teachers who gave the test and who were asked to give their opinion of its value, the constant use of such a scale would improve the composition work in these schools.

The opinion of all the teachers was that "the scale is a quicker and more accurate method of grading themes." An analysis of the teachers' reports shows very clearly the lack of standards as to just what constitutes good written composition work. The prevailing opinion was expressed in the words of one of the teachers: "If a standard for composition is established for each grade, it will be a helpful guide in carrying on the written work."

The use of such an exercise as a basis for determining promotions received general approval. The different judgments can be summarized in the words of one of the teachers who reported as follows: "It seems a fairer and more accurate way of judging a child's ability than the customary examination."

Once teachers realize the necessity for more definite standards in such subjects as written composition and understand the use of such a scale as the Nassau County Supplement and the help derived therefrom, a more scientific procedure is assured. The principal of one of the schools in which the above results were secured reports as follows: "The results of our first use of the Nassau County Supplement to the Hillegas Composition Scale seem entirely favorable.

"A few teachers were at first of the opinion that considerable extra work was involved in the use of the scale. Lack of



familiarity with it caused much more time to be taken with this work than would be necessary after further use. I think that the opinion was changed after the last papers were graded."

The Nassau County Supplement to the Hillegas Scale is, on account of its simplicity, one of the best written composition scales so far available for use by teachers in the elementary schools. It enables them to grade their written compositions more accurately and with greater speed than the present system of marking.

#### WILLING SCALE FOR MEASURING WRITTEN COMPOSITION

**Aim.** — The purpose of this scale is to measure the "Story Value" and also the "Form Value" of the written composition of children in grades four to eight.

**Description and Application of the Scale.** — The scale consists of 8 compositions ranging in value from 20 to 90 in steps ten units apart. "They are all on the same topic, An Exciting Experience. The scale is made up of the material it attempts to measure." By "Story Value" is meant the degree of completeness with which the story in the composition is told; by "Form Value" is meant the "number of mistakes in spelling, punctuation, and syntax per hundred words."

Full instructions are provided for the use of the scale. According to the plan, a class exercise is given on some topic as An Exciting Experience, A Storm, An Accident, etc. Twenty-five minutes are allowed to write the exercise. These compositions are then used as a basis for determining the children's ability in written composition by comparing each composition with the scale and giving it the value of the composition on the scale which it most resembles.

**Scoring the Results.** — The compositions are scored first for the value of the story and second for errors in grammar, spelling, punctuation, and capitalization. If a composition

scores high for the value of the story and low for its form value or vice versa, a score between these two values should be finally agreed upon. According to the author "no paper is marked above 70 which does not have good story value and technical excellence, nor is a paper marked below 40 which does not lack both of these qualities."

Below is given a class record sheet showing the scores from a 6-B grade in a city school system in January, 1920.

TABLE 43. — CLASS RECORD SHEET FOR WILLING'S SCALE FOR MEASURING WRITTEN COMPOSITION

City D School L. P. Grade 6-B  
Teacher A. H. Date Jan. 7, 1920.

ERRORS PER HUNDRED WORDS	20	30	40	50	60	70	80	90	DISTRIBUTION FOR ERRORS
0 to 2.9 . . . . .					1	3	1	4	9
3 to 5.9 . . . . .				2	2	5	2	2	13
6 to 8.9 . . . . .				3	1		1		5
9 to 11.9 . . . . .					1	1	3		5
12 to 14.9 . . . . .			1						1
15 to 17.9 . . . . .				1					1
18 to 20.9 . . . . .									
21 to 23.9 . . . . .									
24 to 26.9 . . . . .									
27 to 29.9 . . . . .									
Above 30 . . . . .									
Distribution for "story value" . . .			1	6	5	9	7	6	

Class medians.

Form value 3.5

Story value 70

This record sheet gives the distribution of the scores of this class according to the story value and the form value of the different compositions. It is read as follows: The papers of nine pupils showed from 0 to 2.9 errors in "spelling, punctuation, and syntax per hundred words." Of this number one

pupil had a story value of 60, three pupils had a story value of 70, etc. The median class scores are form value 3.5 and story value 70.

**Interpreting and Using the Results.**—The following standard scores are available with which comparisons can be made:

TABLE 44. — MEDIAN SCORES FOR WILLING'S COMPOSITION SCALE

GRADE	DENVER		FIVE KANSAS CITIES	
	Story Value	Form Value	Story Value	Form Value
IV . . . . .	32	22	44	12
V . . . . .	43	16	58	10
VI . . . . .	50	14	75	5
VII . . . . .	60	11	77	5
VIII . . . . .	63	10	82	6

By comparing the scores on the above record sheet (Table 43) with the standards (Table 44), it is readily seen that there are two children who are very poor in both story value and form value. Possibly the teacher was acquainted with this fact, but just how much she could not tell unless such a test had been given and standards of attainment for different grades provided. She knows now that one pupil does not exceed the fourth grade standards of the five Kansas cities and the other is below the fourth grade standard in form value and the fifth grade standard in story value in the same cities. She also knows that her class as a whole is above the Denver standards for the eighth grade and below the sixth grade standard in story value and above the eighth grade standard in form value in the five Kansas cities.

An analysis of the errors on the papers from this class will enable the teacher to know the forms which should continue to receive drill. She will also know that steps should be

taken to improve the children in grasping and expressing the story in such a written theme.

The plan of judging the compositions first for the value of the story and second for errors in form enables the teacher to analyze the strength or weakness of her written composition work. This being accomplished, she is able better to stress those elements in which the class is weak and which are essential to good composition.

In the scoring of compositions as to general merit, when errors are taken into consideration, it is possibly better to count only certain types of mistakes. The plan of counting mistakes in the Gary survey is suggestive. The following types of mistakes were observed:

#### IN CAPITALIZATION

1. "Failure to begin a sentence with a capital letter.
2. Failure to capitalize a proper noun.
3. The capitalization of common nouns.

#### IN PUNCTUATION

1. Failure to place a period at the end of a sentence.
2. Failure to place a question mark at the end of a question.
3. Failure to inclose a direct quotation in quotation marks.

#### IN SYNTAX

1. The use of the wrong case form, as 'me and him went.'
2. The use of one word in place of another, as 'it would of (have) been.'
3. Lack of agreement of noun and pronoun, as 'the pieces were about the size — and it break.'
4. Lack of agreement between subject and verb as 'they was.'
5. Use of the wrong tense form, as 'seen' for 'saw.'
6. Use of the double negative.

7. Confusion of dependent and independent clauses, as ' . . . away, but worst thing was that there were not light on the streets and no road but a little path through the wood but I dressed up and took my dog and started off we were not far from home when my dog his name was Rover began to chase after I was a fright to go myself and began. . . . '

8. Omission of words essential to the thought, or the addition of irrelevant words, as 'began to chase after (a rabbit, omitted) I was a fright to go myself.' "

In addition to the diagnosis of the language abilities of her class and a comparison of the results with standards of attainment in other cities, the teacher should also be able to compare her class with other classes in the same school or in other schools of the same city; the supervisor should likewise be able to know the weak and strong spots in her organization.

The following table shows the median class scores for written composition in grades 5 and 6 in three city schools in January, 1920, according to the Willing Composition Scale.

TABLE 45

School	STORY VALUE				FORM VALUE			
	5-B	5-A	6-B	6-A	5-B	5-A	6-B	6-A
1 . . . . .	70	81	80	80	14.2	8	7.3	9.1
2 . . . . .	43	50	65	65	11.1	10.9	10.9	4.0
3 . . . . .	80		70		8.5		3.5	

This table is read as follows: School number one made in story value 70 in 5-B, 81 in 5-A, etc., and in form value, 14.2 in 5-B, 8 in 5-A, etc. The scores for the 6-B grade in school number three are also shown on the class record sheet on Table 43.

A comparison of these results with the standards so far



available shows that in story value school number one stands well; in form value it is above the Denver and below the Kansas Standards. An outstanding fact in this school is that the 6-A grade shows more form errors than the 5-A or 6-B grades, a clear indication of the need for more drill on form. School number three is above school number one and the other standards in both story and form values except in story value for the 6-B grade. School number two is below the other two schools and the Kansas cities in story and form values; it compares fairly well with the Denver standards in story and form values.

Although the differences in scores may be due partly to the scoring of the teachers who were limited in their experience in using the scale, it is evident that the written composition work in school number two is weak. More systematic instruction should be given in training these children to organize their experiences in outline form and to express their ideas orally before undertaking the written work. If this preparatory instruction is given before the children are permitted to write, the written compositions will show marked improvement in the content. Drill exercises on the mistakes in punctuation, spelling and syntax should follow the written themes.

The teachers who gave the composition exercise and scored the papers to secure the results in the above table were provided with all the instructions which accompany the Willing Composition Scale. All of the children wrote on the same theme, "An Exciting Experience." They were allowed twenty-five minutes to do the work, which was done as a class exercise.

All of the teachers reported that "it was easily given." Many of them stated that the papers were difficult to score with the scale on account of the lack of variety in compositions comprising the scale. They agreed that the scale in connection with such an exercise was a fairer method of



determining promotion on written composition than the customary examination.

The two outstanding facts as reported by these teachers are the following: first, the scale applied to such an exercise is a good test of the mechanics and the story value of written composition; and second, the value given to each composition on the scale for spelling, syntax, and punctuation is exceedingly helpful in directing attention to definite strength and weakness.

The Willing Scale for Measuring Written Composition can be used by the teacher with advantage to determine the ability of her children in written composition. By analyzing the written composition work into its "rhetorical and formal qualities," instruction can be more carefully directed so that more satisfactory progress is assured.

#### THE TEACHER'S COMPOSITION SCALE

Probably the greatest good which a group of city teachers can derive from the use of a written composition scale will come from a scale of their own construction out of the material which they intend to measure. For such a purpose, it is recommended that the Nassau County Supplement be used as a basis, although the teachers may find it helpful to make comparison with other scales.

A simple plan for doing this was followed in the construction of the Gary Composition Scale. Compositions were selected and arranged into a scale by matching them with the specimens in the Hillegas scale. The values of the different compositions in the Hillegas Scale were then given to the corresponding compositions selected from the material to be measured.

For the purpose of indicating to teachers the nature of such a scale, the complete Gary Composition Scale is given:

SAMPLE A. VALUE 5

*An Exciting Experience*

I was on a wonderful to now-Yeark and it waos and the journey that I lecekt so mouch butit the fun I head on the trean but the buge-ried in ealmer Nou-Yeaork waos the beast thean. and I belavef is all that I can thank on so godboy.

SAMPLE B. VALUE 15

*The Beacr and a Boy*

Ones their was a littil boy. He went in the weth a gan and a hachet and he saw a beaer. The beaer begain to run after the boy. And their was a oake tree he stared to clamb the tree. The beaer avas after the boy and the beaer stared to clamb the tree. And the beaer clambd and clambd so the boy thart he beter get out of their he fawed a hol in the so the got down in the hol and the beaer con't find the boy. So the beaer went on an lime. The bay clambd out and chot the lime of and the beaer was celed.

SAMPLE C. VALUE 20

Once My nother want to voiter a frinced And she take me a gave me to my aunt thats ny mother want away. And my aunt was chopping wood AndI was playing biut a bucter Came and ny aent want inside The house and I want and take The hatck and was chopping the wood with the hatchek and chop my Finger and my nother wasnot Home so I uat crying and Than ny mother came after There day and I was over ny aunt House I was in bed and ny mother was glad I was Well and I was happy again. And than I have a mark on My frienger it is the left hand Frienger it is the 2ndfrineger and the Than after my mother take me Home and I was a happy girl After But I was till or ny frienger That marke stayed ny mother Was heling my friend and Now I am happy girl as Happy as can be happy.

## SAMPLE D. VALUE 30

*In The Mountains*

As we went to spent our vacation I happen to be right near the mountains I was glad couse I could go and climb just as high as I want for.

So I went with my father and mother we went pvery hiegh it was getting cold already why I think abouve the clouds I want to rich the tops but couldn couse there was ice and it was so sleapry to goe any further se we came baak when we came down there was many more moutains and I disided to go on some others well & ni went it wasnot very hiegh just like others so when nex were clinbing it little to sandy we riched the top alright but whenwe wanted to come down why we couldn mole so we sat down and slide down in that way we couldn get down in that city where we were its too cold in summer sometimes its snowing but this little city was full of trees and mountains.

## SAMPLE E. VALUE 40

*An Exciting Experience*

One day it was very hot and we didn't know what to do.

This was at our Gym. period, so we went over on the lawn and sat under a tree for a while in the shade.

After a while one of the girls said, Let's play something." We all suggested that we would play ghost. We started and played for a long while and then we got tired and we sugested we would play something else so we played leap frog.

We were awfully hot now so we sat down in the shade and rested ourselves.

After a while one of the girls sugested that we would play statue. We had been playing awhile and it was my turn to be be swung around. The girl that was swinging had swung all the other girls and they were pretty heavy then she took me and swung me around fast not thinking of hlow light I was she let me go and I fell on my left wrist. I heard it crack and I thought it was broke, they took me down in the Gym and the Gym teacher bandaged it up. It was not broke, but it was sprained.

I bet there were fifty girls that asked me where I fell what I was doing an how it was done. Well that was the last game that I have played since then.

SAMPLE F. VALUE 50

*An Accident*

We were out at camp No. 133 which is situated near the banks of Deep River. One of the men that stayed at this camp owned a old duck boat which leaked and if you wanted to ride in it you would have to set a certain way ot it would fill with water and soon sink.

My brother saw me paddaling around in it and he decided that he would do it himself.

He weighed about twenty-five lbs. more than me I told him the way to set in it but he would not listen but said that one end was as good as the other.

He jumped in and sat down on the nearest end which was the wrong end and paddaled out into the river. He paddaled down the river for some distance and then turned around to come back. By this time the boat was nearly sinking and we saw him paddeling as fast as he could go to get back to the bank.

But it was of no use the boat began to sink and he tried to get to the right end but in trying to get to the right end he upset the boat and had to swim with all of his clothes on. The water wasn't very cold and he swam all the way to up the bridge pushing the boat with him. He soon was in dry clothes and was none the worse for the accident.

SAMPLE G. VALUE 60

*On the Water*

While at a small lake not far from Gary my small brother had quite an exciting adventure. The lake is quite deep and though my brother can row a boat quite well we never allowed him to go out in a boat by himself.

One day my brother asked mother if he might go down to the

beach. She replied, "Yes, but do not get into the boats." Clouds began to gather and it looked as though we would have a violent electric storm.

It began to get dark and mother thought of Bud so she sent my sister to get him. After a few minutes she came back saying that Bud was not on the beach.

We asked if any one had seen him but no one had. I thought that he might be playing with one of the little boys so I went to a house on the bluff expecting to find him there.

While I was on the porch, which over looks the lake, it began to thunder & lighten & soon the rain came down in torrents. I stood looking out over the lake & and I noticed a boat sway at the other side of the lake.

I ran to my father and told him what I had seen. He hurried to the beach & found that one of the boats was missing. Then he and my sister took a boat and in about twenty minutes they came back with a rather frightened & very much bedraggled little boy.

He, my brother, showed great presence of mind, for when the storm began he was not frightened very much and tried to reach the shore.

You may be sure that he did not dare to go out in a boat alone after that. He told us that he would learn to swim first.

#### SAMPLE H. VALUE 67

##### *My First Morning in Mexico*

We came on a train late at night and had a hard time finding an American hotel in San Louis Potosi. After finding one we went straight to bed, although my mother took plenty of time to lock and prop the door closed. The next morning I woke up early and looked out of the window. The first thing I saw was a rickety old closed up wagon coming down the street drawn by a few very small borros. This strange looking object was the morning street-car, the driver was standing in front blowing a small tin horn for the people to get out of the way.

Some of the Peons were getting breakfast, the mother sat on the street baking tortellias and the family seated in a circle about her



eating as fast as she could bake. The funniest thing about the people eating was that the pigs and dogs ran about the outside of the circle eating the scraps that were thrown them.

Down the street comes a man riding on so small a borro that his feet touch the ground, he is smoking a cigarette and lazily looking about. Behind him walks his wife holding the baby and hiting the borro her husband rides on to make it go. Behind the wife come the children each carrying something. The houses along the street are made of adobe, all the windows have bars on them which make the house look more like a prison. Over the tops of the buildings can be seen the mountains which have nothing but the cenutry plant on them. Some borros are.

SAMPLE I. VALUE 80

With a jar and a somewhat business like jolt the rickty elevator came to a stop on the basement floor. The door swung open and I stepped out into press-room of the Chicago Tribune.

Surrounded by a mass of quivvering steel I was at a loss to know what to do. I was suddenly confronted by a bearded man clothed in ink smeared overall and jumper. A small tight shop cap was set jauntily on one side of his head and a pair of steel grey eyes peered at me thru a rather large pair silver rimmed glasses. He seemed to be saying something to me but the battery of Hoe presses had control of the field and it was only with the greatest difficulty that I could hear what he was trying to tell me.

Beckoning me with an ink stained finger the pressman, for such was the position of this man, piloted me around, under, and even over masses of quivering and roaring steel until we stopped before a press which stood two and one half stories high, a half city block in length and the same in width. The silent pressman paused for a moment to glance with pride at the roaring monsters when he motioned me again and mounting an iron stairway with a brass railing we were soon standing on the top deck of this master press. Two and a half stories below me sixteen large rolls of spotlessly white paper were swiftly unrolling into the press. Wheels within wheels whirled and sang and its very song seemed to say. "I am the Frank A Munsey the worlds largest newspaper press."



The value to teachers in having their own scale even though it be roughly determined is as follows: First, they have a double check on their results by the use of the two scales; second, a composition scale constructed from the material which is to be measured is of more value for purposes of comparison; third, in the construction of such a composition scale, the teachers acquire a viewpoint and an intimate acquaintance with the measurement of written composition which they would not otherwise obtain.

**Other Tests.**—*The Extension of the Hillegas Scale for the Measurement of Quality in English Composition by Young People*, by Dr. E. L. Thorndike, is a general merit scale intended to be used in grades 4 to 12 inclusive. It is made up of 29 compositions grouped according to their quality into 15 units with values ranging from 0 to 95.

Some of the qualities have under them as many as 5 and 6 specimens. On account of the larger number of compositions and a system of marking similar to that which teachers ordinarily use, the scale can be used to a considerable advantage by the teacher.

The *Harvard-Newton Scales* for the Measurement of English Composition are made up of four scales for eighth grade composition, one for each of the four types of composition, narration, description, exposition, and argumentation. The scales were constructed by Dr. Frank W. Ballou with the aid of eighth grade teachers in Boston. The compositions were written by eighth grade students. Each scale is made up of six compositions with values ranging roughly from 40% to 95%. An important feature of each scale is a notation of the merits and the defects of each composition and a comparison with the compositions above and below it.

*A Scale for Measuring the General Merit of English Composition*, by F. S. Breed and F. W. Frostic was constructed from compositions written by sixth grade pupils. It is a general merit scale.

A *Punctuation Scale* has been constructed by Dr. Daniel Starch for the purpose of determining a child's ability to punctuate. It is made up of a number of sentences arranged in a series of 10 steps which the child is to punctuate. These sentences increase in difficulty with each step. Tentative standards of attainment have been formulated for the seventh and eighth grades.

A *Copying Test* was used by a group of Boston teachers to determine the degree of accuracy with which pupils copy. It is intended primarily for the grammar grades and the high school. Such errors are noted as occur in the following: Spelling, capitalization, omitted words, and added words.

## BIBLIOGRAPHY

### REFERENCES

- Hillegas, Milo B., "A Scale for the Measurement of Quality in English Composition by Young People," *Teachers College Record*, September, 1912, Bureau of Publications, Teachers College, Columbia University, New York City.
- Monroe, Walter S., "Measuring the Results of Teaching," Houghton Mifflin Company. "Existing Tests and Standards," Seventeenth Year-book, 1918, The Public School Publishing Co., Bloomington, Illinois.
- Monroe, W. S., De Voss, James, and Kelley, Frederick J., "Educational Tests and Measurements," Houghton Mifflin Company.
- Breed, F. S., and Frostic, F. W., "A Scale for Measuring the General Merit of English Composition in the Sixth Grade," *Elementary School Journal*, Vol. 17, pp. 307-325.
- Denver Survey, Denver, Colorado.
- Grand Rapids Survey, Grand Rapids, Michigan.
- Gary Survey, "Measurement of Classroom Productions," pp. 216-262, and Appendix A, pp. 416-437, General Educational Board, 61 Broadway, New York City.

### SCALES

- Breed, F. S., and Frostic, F. W., "A Scale for Measuring the General Merit of English Composition in the Sixth Grade," *The Elementary School Journal*, The University of Chicago Press, Chicago, Illinois.

- Ballou, F. W., "Harvard-Newton Composition," Harvard University Press, Cambridge, Massachusetts. Price, \$.50.
- Hillegas, M. B. "Hillegas Composition Scale." Bureau of Publications, Teachers College, Columbia University, New York City. Price, Chart, 6"  $\times$  20", \$.03 by mail; in quantities of 25 or more, \$.02 per copy, postage extra.
- Trabue, M. L. "Nassau County Supplement to the Hillegas Scale." Bureau of Publications, Teachers College, Columbia University, New York City. Price, \$.08 by mail; in quantities \$.05 per copy, postage extra.
- Starch, Daniel, "Starch's Punctuation Scale," University of Wisconsin. Price, \$.45 per hundred. Directions, \$.01 per copy.
- Thorndike, E. L., "Thorndike Extension of the Hillegas Scale for the Measurement of Quality in English Composition by Young People." Bureau of Publications, Teachers College, Columbia University, New York City. Price, single copy, \$.08 by mail; in quantities, \$.05 per copy, postage extra.
- Willing, W. H., "Willing Scale for the Measurement of Written Composition. Bureau of Coöperative Research, Indiana University, Bloomington. Price, \$.08 per copy; five or more, \$.05 per copy postage extra, \$.01 per copy.

## CHAPTER VII

### THE MEASUREMENT OF DRAWING

**How is Drawing Now Measured?** — To the teacher of artistic temperament who objects that accurate grading of work in drawing is impossible, that much work is dependent upon sentiment, is intangible, and should be left unmeasured so far as the pupil is concerned, the answer is that measurement in drawing is constantly being made through "grades" or "marks" which are given to children. Grades ranging from 100% down to 60% or from A to F are given year after year, term after term, and in some school systems month after month. The effort, therefore, to construct a scale that will make possible more accurate measurement of drawing should be welcomed by every teacher of drawing. Notwithstanding millions of measurements or marks each year, there are few teachers who themselves have any adequate standard for grading which recognizes definite differences in ability between the pupils of one grade and another, or which makes possible any accurate comparison of the drawing work among buildings. The old marking or grading system has given us no accurate method for measuring products in drawing. Pupils themselves protest again and again against grades received that are based upon the mere whim of the teacher. Or perchance the teacher was out late the night before, or had a slight case of indigestion, and the pupil's grades are cut accordingly. This is but another case of a badly felt need for objective standards in grading.

**The Scale and Some Uses of It.** — In 1913, Dr. E. L. Thorndike constructed a tentative scale for the measurement of

general merit in drawing. The use of such a scale in the schoolroom is not only advantageous to the teacher, supervisor, and superintendent, but it is appreciated by the pupil as an instrument of fairness and democracy. With a definite scale available for comparisons, pupils themselves will arrive, after some practice, at reasonable accuracy in judgment and they will be able to see clearly their own progress from a lower to a higher stage of ability. Dr. Thorndike's scale, while tentative and while making no claim for perfection, does quite surely make possible the comparison of a pupil's work at one time with his work at another time. It permits comparison with reasonable accuracy between the work of one pupil and another pupil or between one room or group of children and another. It permits comparison of work in one city with the work in another city. It enables supervisors or superintendents to determine the relative results from one hundred minutes per week in one part of the city over against fifty minutes per week in another part of the city and so makes it possible to determine the time allotment for drawing on a reasonably intelligent basis. It should permit also comparison of one method of instruction with another when conditions are properly arranged and conclusions properly evaluated. None of these results were possible except in a very general and indefinite way under the old marking system. Under the old system if one hundred pupils in an entire city had a grade of 95% or above, and if such pupils were scattered in as many different rooms of the city, no one could infer that these one hundred pupils had the same ability or proficiency in drawing. On the other hand if as many as four pupils in the sixth grade of a large city system reached a proficiency equal to merit 17 of the Thorndike scale there would be little doubt on the part of any one familiar with the system that these pupils were fairly equal in ability and that they showed ability very much above the average. In fact merit 17 is such a degree of excellence as would be reached





genera  
school  
visor,  
as an i  
scale a  
after s  
they v  
lower  
while  
does c  
work i  
comp s  
one p  
of ch i  
one c i  
or su f  
hund i  
fifty i  
make i  
on a  
comp i  
condi i  
evalu  
very  
system  
entire  
scatte  
could  
or pr  
four j  
profic  
woul  
syste  
they  
merit

by only one pupil out of 50,000, ages 6 to 15, and to discover a few pupils in a large city system who had superior ability can readily be recognized as information of unusual value.

**Requirements of a Scale.** — A scale in drawing as in any other subject should permit accurate measurement of differences in degree of merit. To weigh out sugar in five, ten, or fifteen-pound packages is comparatively simple as the method of weighing has been perfected to a very high degree. No such degree of accuracy should be expected from the use of the drawing scale. The samples of merit on the scale as arranged cannot be handled in the same definite and accurate manner as the pound weight which the grocer places upon the left-hand side of the scale to balance a pound of sugar placed on the right-hand side, but a little practice will enable the teacher to rank drawings with a fair degree of accuracy. This is the proper function of a scale. As the teacher becomes more familiar with the scale, and increases her experience in its use she will acquire more and more the weight-accuracy secured in handling sugar, but it is reasonable to expect that she will never quite reach sugar-weighing accuracy.

An explanation of the unit of the Thorndike scale for drawing will make clear that absolute accuracy is hardly to be expected. The formation of the scale is built upon the basis of equal differences in judgment and the unit chosen for this particular scale is: *the difference of merit in children's drawings which 75% of artists, teachers of art, and intelligent judges generally can distinguish, and which 25% of them fail to distinguish.* It will be observed that the unit itself allows for 25% error in judgment by those who are competent to judge. If the teacher makes mistakes in equal ratio, one in every four, the teacher still improves her judging by the use of the scale. It should be noted, however, that the units in the Thorndike scale are sufficiently large that the error in placing a drawing is seldom as great as a full step and that the usual

variation of a judge does not exceed half a merit. Dr. Thorndike has explained this point as follows: "If the same judge should so rate a thousand drawings, and then, putting these ratings aside, rate the thousand over again he would vary often by more than half a 'merit' from his previous judgments." This means that for a high degree of accuracy a teacher should rate drawings several times and then take the average, or in a contest of unusual importance, the drawing should be rated by several judges and the average of the judgments taken. The necessity of a large number of rankings is greatly reduced by the use of a drawing scale. This was demonstrated by an actual test. Ten teachers measured the merit of a drawing by the use of the scale and varied only 4 points all told. Ten other teachers measured the merit of the same drawing without the scale. They were instructed to grade the drawings from 0 to 17. They showed a variation of 14 points, or nearly 4 times as much variation without the scale as with it.

**How the Scale was Derived.** — The scale as derived by Dr. Thorndike is based upon a preliminary study of 45 drawings and a more intensive study of 15 drawings. The 15 drawings were submitted to 376 judges, 60 of them being artists of sufficient distinction to be listed in "Who's Who in America," 80 being supervisors or teachers of art, and 236 being students of education and psychology. The judges were asked to rank the 15 drawings according to merit. These 376 judgments of merit permitted Dr. Thorndike to determine the scale steps on the basis of differences in judgment of merit. The table which follows herewith summarizes the judgments with reference to the relative merits of the 14 drawings finally used in the scale.

With this table as a basis and with a unit defined as the difference of merit recognized by 75% of the judges, Dr. Thorndike was able by the use of statistical methods to assign a definite value to each of the 14 drawings of the scale.

TABLE 46

94.85	per cent of the judges rated <i>b</i>	as better than <i>a</i>
84.5	per cent of the judges rated <i>c</i>	as better than <i>b</i>
88.45	per cent of the judges rated <i>d</i>	as better than <i>c</i>
69.5	per cent of the judges rated <i>e</i>	as better than <i>d</i>
82.55	per cent of the judges rated <i>f</i>	as better than <i>e</i>
69.7	per cent of the judges rated <i>g</i>	as better than <i>f</i>
89.4	per cent of the judges rated <i>h</i>	as better than <i>g</i>
81.75	per cent of the judges rated <i>i</i>	as better than <i>h</i>
70.	per cent of the judges rated <i>j</i>	as better than <i>i</i>
73.35	per cent of the judges rated <i>k</i>	as better than <i>j</i>
72.5	per cent of the judges rated <i>l</i>	as better than <i>k</i>
86.5	per cent of the judges rated <i>m</i>	as better than <i>l</i>
74.2	per cent of the judges rated <i>n</i>	as better than <i>m</i>

It is scarcely to be assumed that a unit in one part of the scale is exactly equal to a unit in another part of the scale. It is, however, safe to assume that the scale is entirely on a ranking basis, that is, that the drawings are arranged in the order of merit. It is also safe to assume that the units are approximately the same in one part of the scale as in another and that progress of one unit in any part of the scale is approximately equal to similar progress in another part of the scale. Dr. Thorndike gives warning, however, that the exact numeral relations of the scale should not be used too freely in scientific quantitative studies of achievement and improvement in drawing. For the practical purposes of measuring drawing work in the public schools, however, no harm will result from taking the scale at its face value.

While it is unnecessary to attempt to understand in detail Dr. Thorndike's method of determining the values of the drawings in the scale, an idea of the procedure may be easily grasped by considering the extremes of the situation.<sup>1</sup> If drawing *x* were judged as better than drawing *y* by exactly 50% of a large group of competent judges, it is

<sup>1</sup> Students interested in fully understanding the procedure are referred to Thorndike's "Mental and Social Measurements," Chap. VIII.

apparent that just as many judges thought  $x$  better than  $y$  as thought  $y$  better than  $x$ . If the judges were sufficiently competent and numerous, we are justified under the conditions in assuming that drawings  $x$  and  $y$  are equal in merit. On the other hand if 100% of a large group of competent judges ranked drawing  $y$  as better than  $x$  we would only know that  $y$  was "vastly superior" or "so superior as to be in an entirely different class." Now, manifestly, there are all grades of difference ranging from a 50-50 judgment of equality to a 0-100 judgment of superiority, and some place in between these two would be found the degree of superiority that is recognized by only 75% of the judges. This is the unit of merit of the Thorndike Drawing Scale. By calling the first drawing (drawing  $a$ ) in the scale *zero*, and remembering from the table above that 94.85% of the judges rated drawing  $b$  better than drawing  $a$ , the mathematics of the case gives 2.4 as the merit value of  $b$ . In like manner the other values of the scale were determined, — 3.9, 5.7, 6.5, 7.8, 8.6, 10.5, 11.8, 12.6, 13.5, 14.4, 16.0, and 17.0. The last two steps in the scale are almost exactly 1 merit apart, as  $n$  was judged better than  $m$  by 74.2% of the judges. So if  $m$  has a value of 16.0,  $n$  should have the value 17.0. It thus appears that such an intangible situation as differences in judgment may be made the basis of mathematical calculation for the formation of a definitely evaluated scale.

**Limitations of the Scale.** — The chief limitation of the scale is that it is based upon *general* merit in drawing and is, therefore, not analytical of different types of drawing. In fact the scale combines various types of drawing into a single scale and thus relates to comparisons of materials that in many respects are not comparable. In time we should have scales developed that will deal with merit in particular lines rather than general merit. For instance the following possible unit scales will be recognized by a teacher of art and drawing:



- a. Pencil drawings.
- b. Charcoal work.
- c. Sepia originals.
- d. Sepia reproductions.
- e. Water color.
- f. Mechanical drawings.
- g. Maps and informational designs.

The teacher may even make divisions along entirely different lines, resulting in such divisions as decorative, photographic, informational, representative, etc.; or even along such simple lines as types of subjects, as, trees, buildings, fruit, animals, etc.

The type of art work carried on in different cities varies considerably. The supervisor of art and drawing will doubtless find it especially helpful to use the Thorndike drawing scale for grading her work until she can supplement it by other scales, made by ranking and evaluating the work of her own pupils. Scales resulting in this way will be more easily understood by grade teachers and pupils, and they will also be more fully appreciated. The most useful scale for lower grade work will doubtless be a scale of merit for products in water color. Comparison with the Thorndike drawing scale will be the easiest and simplest method of procedure in making such a scale. At first the supervisor may have difficulty in filling in all of the steps on the scale. It may take the larger part of a year to complete this scale and others needed to cover the various phases of art work undertaken in the particular city, but all the while she will be collecting samples, comparing, rejecting, replacing, and getting her own ideas more and more definite. The result will be that by the time such scales are completed their value will be apparent and their usefulness unquestioned. Scales thus made up from the work of the pupils in a school system will be more useful in many respects and will accomplish some purposes not possible with the Thorndike scale, even though they are



much less accurate and not at all scientifically constructed. They will result in constant reference to the Thorndike scale and in a much better understanding of it.

With reference to the extension and refinement of his own drawing scale, Dr. Thorndike makes the following comment: "These limitations can be remedied in part, though we may not expect to measure the merit of a child's drawing as easily and accurately as the weight of his body. The scale can be extended in scope to include specimens of maps, mechanical drawings, decorative drawings and designs of various sorts, drawings of specified objects, and the like; and with proper methods of investigation and enough labor, specimens can be found of these several sorts of drawings of exactly 1 'merit,'  $1\frac{1}{2}$  'merits,' 2 'merits,' and so on up to 17 'merits,' which is as far as a scale for children's drawings needs to extend. The labor involved is, however, very, very great. In order to get one specimen proved to be between .99 and 1.01, it would probably be necessary to collect with great care at least fifty drawings by very stupid children, to have them measured by at least 100 judges in comparison with the scale as it now is, then to select 20 of them to be so measured by 200 more judges, and then to select ten of these to be measured by 300 more judges, and finally to have the two or three or four of these that were between .99 and 1.01, by the opinion of the 600 judgments so far, measured by 400 more judges. If the present scale were not at hand as a basis, the labor would be much greater. The improvement of the scale in these respects must then be a gradual achievement of several years."

**Grade Standards.**—The fixing of grade standards in drawing or the various phases of art work will gradually follow an extensive use of scales with definitely evaluated units. The first step in fixing grade standards will be to agree upon a set of rules or instructions for taking samples of the work of pupils. These instructions will doubtless need

to indicate a range of subjects or problems within the ability of the various grades tested and then fix time limits for the tests. However, this is only a guess as to how the instructions should be drawn. It may be that there should be no time limit, or that the best work of a pupil for a month or a term should be taken for grading. The art supervisor who works out satisfactory instructions and fixes grade standards will be performing a worth while educational service.

Childs reports<sup>1</sup> a preliminary use of the Thorndike scale during the winter of 1914. A total of 2177 pupils in two school systems was involved in the test. For the purposes of the test, a new scale was constructed by including all the human figure samples of the original Thorndike scale, and adding snow scene samples, thus making a scale with a total of 16 human figure and snow scene samples. In administering the test, a time limit of ten minutes was observed and the subject was assigned as follows: "Scene or picture with snow on the ground and boys or girls doing something, as snow-balling, coasting," etc.

The following median scores resulted from the test:

TABLE 47

GRADE . .	1B	1A	2B	2A	3B	3A	4B	4A	
Median . .	5.5(?)	6.1	5.9	7.5	7.9	7.6	7.5	8.3	
GRADE . .	5B	5A	6B	6A	7B	7A	8B	8A	All
Median . .	8.2	8.4	8.5	9.7	8.7	9.3	8.8	9.5	8.09

Childs thinks that the above medians should not be taken as norms for grade performance, but that norms, if based upon this study, would be by grade groups as follows:

<sup>1</sup> See Bibliography at close of chapter.

TABLE 48

GROUP	GRADE	CHARACTERISTIC SCORE
I . . . . .	1B	4.5(?), 5.5
II . . . . .	1A-2B	5.5, 6.5
III . . . . .	2A-4B	7.2, 8.0
IV . . . . .	4A-6B	8.0, 8.6
V . . . . .	6A-8A	8.6, 9.3

This study, while brief and preliminary in nature, confirms Dr. Thorndike's suggestion that differentiated scales are desirable. The "human figure, snow scene scale" used in this study was much more satisfactory than the Thorndike general scale. The study also confirms the findings of Barnes, Ludens, Burk, and Gotze, that children show a plateau of non-development in drawing from the age of nine or ten on to adolescence. It leads, therefore, to the question as to whether the time schedule in drawing should not be greatly reduced for most children from ages 10 to 14. There is little doubt but that extensive use of objective standards for grading products in drawing would do for that subject what it has done for other subjects. It would lead to a better selection of subject matter, improved methods of instruction, a revision of time schedules, and provision for handling children according to capacity and needs.

**Using the Thorndike Scale.** — The drawing scale is used in the same manner as the writing scale. The specimen of the pupil's work is moved back and forth on the scale until the scale specimen to which it corresponds most nearly in general merit is found. This gives the value of the pupil's work in terms of the scale, and the scale value should be placed as the "grade" on the pupil's work, if a "grade" is desired. For example, if the pupil's drawing corresponds in general merit to the sixth specimen on the scale, merit 7.8, then the pupil's grade is 7.8.

When first beginning the use of the scale there will be a tendency to place the sample on the scale according to the kind of drawing, rather than general merit. This is a difficulty which is soon overcome by practice in the use of the drawing scale just as it is in the use of the writing scale. The scale is constructed on the basis of general merit, and the specimen is to be located on the scale on the basis of general merit regardless of the kind of drawing. Pupils have difficulty on this point when the scale is first used, but it is surprising how quickly pupils as well as teachers advance to the stage of recognizing general merit in the use of the scale.

#### BIBLIOGRAPHY

1. Thorndike, E. L., "The Measurement of Achievement in Drawing," *Teachers College Record*, November, 1913. Copies of the Thorndike Drawing Scale may be secured from the Bureau of Publications, Teachers College, West 120th Street, New York City.
2. Childs, H. G., "Measurement of Drawing Ability, etc.," *Journal of Educational Psychology*, 6: 391-408, September, 1915.
3. Whitford, W. G., "Empirical Study of Pupil Ability in Public School Art Courses," *Elementary School Journal*, 20: 33-46, September, 1919, and 20: 95-105, October, 1919. For copies of the Whitford Scale, address the Prang Company, Chicago.

## CHAPTER VIII

### THE MEASUREMENT OF OTHER GRADE SUBJECTS

THE tool subjects of the grades are being measured with success and with beneficial results on teaching and curricula making. Can the content subjects — such as history, geography, physiology, literature, nature study, and elementary science — be measured with equal success and equally beneficial results? The answer is that many attempts are being made, that success has not been attained, and that final success is still in doubt. A scientific test or scale for grading a subject is merely a reasonable examination which has been carefully graded and evaluated, *i.e.* standardized. Any fixed or rigid examination scheme tends always to formalize the teaching of a subject. For the formal phases of the tool subjects this is desirable, assuming good teachers and provision for adequate motive. But can we formalize the teaching of a content subject without undesirable results, or can we apply standard tests to the more formal informational phases of such a subject without its resulting in misplaced emphasis by many teachers, a large majority of them? It is very doubtful. At any rate, it remains an open question.

Standardized tests<sup>1</sup> have been devised in United States history, geography, language, and grammar.

**History.** — At least seven standardized tests are available in United States history. Five of these test information only; namely, those by Bell and McCullum, Harlan, Starch, Davis, and Raynor. One test, by Van Wagenen, is devised

<sup>1</sup> See Bibliography at the close of the chapter for references and list of available tests.



to test thought and character-judgment as well as information. Barr is working on a set of diagnostic tests in United States history, covering information, thought, reasoning, and judgment. The form of the history test varies. Harlan, Starch, and Raynor use the completion test form. Every test has made an effort to make the questions simple and definite so as to secure answers that can be graded as either right or wrong, thus simplifying and standardizing the grading. In the tests dealing with thought, reasoning, or judgment, this can be done to a limited extent only.

The *Bell and McCullum test* is one of the first devised for testing history and is a good illustration of the informational type. It was first made available in 1917, but apparently it has not been extensively used. The test consists of seven parts, as follows:

I. Give the reason for the historic importance of each of ten representative dates (Dates — Events). II. Indicate for what each of ten prominent men was celebrated (Men — Events). III. Mention the name of the man prominently connected with each of ten historic events (Events — Men). IV. Define in a short sentence each of ten historic terms (Historic Terms). V. Make a list of all the political parties that have arisen in the United States since the Revolution, and state one principle advocated by each (Political Parties). VI. Indicate the great divisions or epochs of United States history (Divisions of History). VII. On an outline map of the United States (supplied) draw the land boundaries of the United States at the close of the Revolution, and indicate the different acquisitions of territory since that date (Map Study). The questions were as follows:

I. Dates — Events. (Four minutes.)

- |          |           |
|----------|-----------|
| 1. 1861. | 6. 1619.  |
| 2. 1789. | 7. 1783.  |
| 3. 1620. | 8. 1492.  |
| 4. 1565. | 9. 1776.  |
| 5. 1898. | 10. 1846. |
| 0        |           |



- II. Men — Events. (Five minutes.)
  - 1. John Burgoyne.
  - 2. Alexander Hamilton.
  - 3. Jefferson Davis.
  - 4. Walter Raleigh.
  - 5. John C. Calhoun.
  - 6. Cyrus H. McCormick.
  - 7. George Dewey.
  - 8. Sam Houston.
  - 9. Roger Williams.
  - 10. James Oglethorpe.
- III. Events — Men. (Three minutes.)
  - 1. Captured Quebec during French and Indian War.
  - 2. Discovered the North Pole.
  - 3. Wrote the Declaration of Independence.
  - 4. Invented the telephone.
  - 5. Brought about the Missouri Compromise.
  - 6. Captured the City of Mexico during the Mexican War.
  - 7. Founded the Colony of Maryland.
  - 8. Made a great speech against the English Stamp Tax.
  - 9. Was President of the United States during the Civil War.
  - 10. Vetoed the re-chartering of the United States Bank.
- IV. Historic Terms. (Seven minutes.)
  - 1. Second Continental Congress.
  - 2. Lewis and Clark Expedition.
  - 3. Articles of Confederation.
  - 4. Sherman Anti-trust Law.
  - 5. Monroe Doctrine.
  - 6. Fugitive Slave Law.
  - 7. Dred Scott Decision.
  - 8. Alien and Sedition Laws.
  - 9. Nullification Ordinance of South Carolina.
  - 10. Emancipation Proclamation.
- V. Political Parties. (Five minutes.)
- VI. Divisions of United States History. (Five minutes.)
- VII. Map Study. (Five minutes.)

The tests are easily administered. There should be further specific directions as to scoring the separate questions. The tests were originally given and standardized on the basis of the answers of students selected from the Texas normal schools and the University of Texas. There were 523 students from grades six and seven, 668 high school students, 207 normal school students, and 75 students from the University of Texas. No attempt has been made to fix grade standards. The test was used originally in order to study the question, "What will a carefully constructed information test in United States History reveal regarding individual, sex, and school differences?"

Doubtless the most valuable purpose that can be served by this scale is that of the study of the effectiveness of various methods in fixing traditional facts in the minds of the children. The test is one of the most valuable of existing tests on the old type of history, which has for its end the mastery of the facts in the traditional course. It is doubtful if the test affords even a comprehensive review of old chronological history or if the details of each test are well selected. Under the Dates—Men test, 1846 would not be selected as an important date outside of Texas. It is doubtful if 1565 is one of the important dates in United States history. Under the Men—Events test, it is very doubtful if the ten most important men in our history are mentioned. There is a tendency, throughout the entire series of tests, to place as much emphasis on the earlier phases of United States History as upon the later. An examination of these tests gives rise to the question as to whether or not they will perform any desirable service in the hands of teachers for examination purposes. The same question is properly raised with reference to the Harlan tests, the Starch tests, the Davis tests, and the Raynor tests. They seem to miss the fundamental purpose of the review or examination as an instrument of teaching, and their tendency is to place emphasis upon the phases of history that are less important.

FIG. 11

## VAN WAGENEN AMERICAN HISTORY SCALES

## INFORMATION SCALE A

Name. .... Sex. .... Grade. .... School. ....  
 When was your last birthday? .... How old were you? .... Date. ....

1. What people did Columbus find in America?	2. Name any American general.	3. In what did the Indians live?
4. Who was President of the United States during the Civil War?	5. By what people was our Thanksgiving Day custom started?	6. With what country did the United States have war in 1898?
7. Name any man besides Columbus who made early explorations in America.	8. In honor of what event do we celebrate the Fourth of July?	9. What were the two chief occupations of the Indian men?
<p>10. Arrange these events in the order in which they occurred by putting a "1" before the event that occurred first, a "2" before the event that occurred second, and so on until you have put a "5" before the event that occurred last.</p> <p>..... Struggle between the French and the English for control in America.</p> <p>..... Rise and growth of the United States as a nation.</p> <p>..... Discovery of America.</p> <p>..... Settlement of America by European nations.</p> <p>..... Struggle of the American colonies against European control.</p>		
<p>11. In what war was the battle of Gettysburg fought?</p> <p>The battle of Trenton?</p> <p>The battle of Lake Erie?</p>	<p>12. What was Henry Hudson looking for when he sailed up the Hudson river?</p> <p>13. Who was President of the United States when Louisiana was purchased?</p>	<p>14. What were the first four European countries to make settlements in America?</p>
<p>15. Who was the British general in each of these battles:</p> <p>Battle of Saratoga?</p> <p>Battle of Yorktown?</p>	<p>16. During what war did iron war vessels first come into use?</p> <p>17. What group of Indian tribes lived in the western part of New York State?</p>	<p>18. What important means of communication were invented and put into use between 1835 and 1845?</p> <p>Between 1870 and 1880?</p> <p>Between 1895 and 1910?</p>

The *Van Wagenen tests*<sup>1</sup> are referred to as scales. There is an information scale, a thought scale, and a character judging scale. It is doubtful if the term scale is properly applied. They will be referred to here as tests. The information test is more extensive than the Bell and McCullum test. It consists of 32 questions, some of which have several parts that are practically equal to additional questions. Figure 11 shows a section of the scale, including the first 18 questions. It is quite evident that the author, in these questions, is attempting to get a good sampling of the individual responses of the pupils on history information. How fully and to what advantage they can be used by the individual teachers yet remains to be seen. For the superintendent who desires comparison among schools or teachers, or for the educational expert, who desires to survey an entire school system, they will afford comparisons which should form the basis of valuable inferences. The effects upon the curriculum of frequent uses of the test will need to be watched carefully and properly guarded.

The thought test is significant in that it recognizes the importance of thought or content considerations in the study of history. It is at least to be commended as a first attempt at attacking this difficult phase of history work. Questions 1, 2, 3, 7, 13, 19, and 22, which follow herewith, are illustrative of the questions in the thought test.

1. Before the steamboats were made people used to travel on the ocean in sailboats. Steamboats were not made until a long, long time after the European people came to make their homes in America.

How do you think these early European settlers came to America?

2. A little before the year 1500 the people of Europe were anxious to find a new way to get to India. Some people thought that India might be reached by sailing westward across the At-

<sup>1</sup> See Bibliography at close of chapter.

lantic Ocean. Columbus was one of these people. It was at this time that Columbus found America.

What do you think Columbus was looking for when he found America?

3. A hundred years ago it took a letter several days to go from New York to Boston. To-day it takes only a few hours.

Why do you think it took letters so much longer to go from New York to Boston 100 years ago than it does to-day?

7. In 1829-30, it took over 160 hours of work to raise 50 bushels of wheat; in 1895-96, it took less than seven and a half hours of work to raise the same amount.

How can you account for the difference?

13. In 1660, the English Parliament passed the restrictions that certain colonial products, called enumerated articles, including sugar, tobacco, dyewoods and indigo, should be shipped from America only to England or to other English colonies.

In 1663, an act of Parliament provided that all goods brought to the colonies must come from or through English ports.

What do you think was the purpose of the English in thus seeking to regulate the trade of the colonies?

19. At the outbreak of the Civil War there were comparatively few factories for spinning and weaving of cloth in the South. They could no longer get cloth from the North and the Northern blockade shut it out from England. Besides they had little machinery and no means of making machinery for spinning and weaving.

In such a crisis how do you think the people of the South obtained the cloth necessary for clothing?

22. At the close of the Revolutionary War many of the people in America were driven from their homes by official acts of a new state government, their property was taken, and they were deprived of the right to vote or to hold public offices.

How can you account for such action?

No attempt will be made to give a critical evaluation of this test. That must be determined by its more extensive use. It may be noted in passing, however, that 13 of the 22 questions in the thought scale relate to history preceding 1812.



Studies which have been made of the relative importance of facts in American history as indicated by present social usage clearly point to the fact that the author of this test has greatly misplaced the emphasis. In fact, there is little evidence throughout the entire test that the author has in mind that history can be used in solving present-day problems.

The character-judging test consists of fifteen questions dealing respectively with the following topics: (1) white man's response to Indian treachery, (2) Nathaniel Hale, (3) John Quincy Adams's refusal to remove a political opponent from office, (4) John Quincy Adams and the right of petition, (5) an Indian father's love for his son, (6) Fletcher and the Earl of Belmont as governors of the New York Province, 1692-1698, (7) English Colonial soldiers, against the Indians in Massachusetts, 1724, (8) Secretary Stanton's behavior in tearing up a decree from President Johnson, (9) Indian Warfare, (10) Indian Warfare, (11) Indian Warfare, (12) Parliamentary retort, (13) St. Clair and Butler against the Northwestern Indians, (14) Political prejudice, (15) Difference between Lieut. Derby and Secretary of War Davis during President Pierce's administration.

It will be observed that 7 of these 15 questions deal with Indians or Indian warfare in some form. One deals with colonial government, at least two with the question of political prejudice, and the latest date of any of the events is the one referring to Secretary Stanton during the administration of President Johnson. In view of this analysis, one may properly doubt the adequacy of the questions for testing character-judgment in history, particularly on a basis of present utility. The characters are too far removed. The appeal is not in any case strongly motivated. The examination, therefore, with this set of questions is sure to be largely a formal matter so far as the children, or even the teacher, are concerned.



Questions 1, 8, and 12 are quoted herewith.

1. In 1772, there was a frontier wedding. The guests had come from many miles. After a night of rough merriment and dancing the guests lay down to sleep under the roof of their host or in the near-by barns and sheds. When morning came two of their horses were missing. Not doubting that they had strayed away, three of the young men started out to find them. Soon several gunshots were heard and the three young men did not return. Believing that it was a small scalping party of Indians eight or ten more mounted the horses that stood saddled before the house and galloped across the fields in the direction of the firing; while others ran to cut off the enemy's retreat.

Draw a line under the three of the following words which you think best describe the action of these white men.

indifferent	cowardly	cautious	polite	brave
courageous	spiteful	fearful	daring	timid

8. General Grant had been very positive in demanding that all officers of the Confederate army should enjoy their liberty. Among those who had been imprisoned by order of the Secretary of War, Edwin M. Stanton, was General Clement C. Clay, an ex-United States senator from Alabama. He was taken ill in prison with asthma, and his wife came to Washington to solicit his release. She went to President Johnson, and he gave her the necessary order, which she took back to Secretary Stanton. Stanton read the order, and, looking her in the face, tore it up without a word and pitched it into his waste-basket. The lady arose and retired without speaking; nor did Stanton speak to her.

Draw a line under the three of the following words which you think best describe this action of Secretary Stanton.

cautious	tactful	callous	generous	courteous
thoughtful	sympathetic	rude	insolent	considerate

12. General Smyth was remarkable for long, prosy, interminable speeches in the House of Representatives. On one occasion, in the committee of the whole, after having wearied the patience of the members more than usual, he said to Mr. Clay, who sat

near him, in a low voice, while he was pausing for a new start, "You speak for the present generation; I speak for posterity," — "Yes," replied Mr. Clay, "and you seem resolved to continue speaking till your audience arrives."

Draw a line under the three of the following words which you think best describe this action of Henry Clay.

kind	bitter	sarcastic	generous	cautious
humorous	ignoble	abusive	sympathetic	ready-witted

The makers of history tests, thus far, have failed to comprehend the true purpose of a review or examination lesson. Such a lesson should give a new view, should secure a reorganization of subject matter, and should provide for use or application. It should, in the best sense of the word, re-enforce the true purposes of the original teaching of the subject. In the case of history, if it is to serve the purpose of civic efficiency on the basis of social utility, much greater emphasis must be placed upon the organization of the material in the form of large motivated problems which look forward to present-day applications. The time spent on the modern periods of history must be greatly increased. The provision for carrying over and applying to present-day problems must be made on a more adequate basis. A review or examination lesson should serve all of these purposes. To date, no test has been standardized which does accomplish these purposes. If standard tests are to be of service in history, they must be so constructed as to effect the desirable basic aims and outcomes of the subject. When this purpose has been served, the other subsidiary purposes, such as distribution of the children according to ability, a diagnosis of classroom results, and a more accurate grading of pupils, will follow.

**Diagnostic Tests in History.** — Dr. Truman L. Kelly, in an experimental study of the analysis and prediction of ability of high school pupils,<sup>1</sup> has included a history test. This has not been developed and used sufficiently to indicate

<sup>1</sup> See Bibliography at the close of the chapter.

its value, but there is, in this use of a test in history, a suggestion of possibilities which needs further attention. A test which is used merely to discover ability, in order to properly advise students to continue further work in the line, or to discover lack of ability, in order to advise students to discontinue work, — this is a use of the test which is less likely to formalize a content subject and which, when properly understood, has connected with it no undesirable results.

**Geography.** — The tests available in geography at the present time are tests of the formal informational phases of the subject. The Starch tests<sup>1</sup> cover the elements of five geography texts, and have been arranged in five parallel tests of equal value. They are in the form of mutilated sentences. They are of limited value for the reasons given in the introductory paragraphs of this chapter and under the discussion of history.

**The Hahn-Lackey Geography Scale.** — The Hahn-Lackey geography scale<sup>1</sup> is an illustration of the application of scientific procedure on an extensive plan, the result being a scale involving both fact and thought questions developed on the plan of the Ayres spelling scale. The scale consists of about 200 questions, graded for difficulty for the fourth, fifth, sixth, seventh, and eighth grades. The questions are based upon *textbooks* and cover the common subject matter of six recent texts. The scale, with complete instructions for grading each question, can be secured from the authors. Figure 12 shows typical data from the scale. A careful study of this section will doubtless convince any reader that the larger purposes of geography will not be furthered by the use of the test. While an attempt is made to involve thought questions, it will be seen that even these deal chiefly with fact or informational phases of the subject. It is doubtful, therefore, if the general use of this scale will be an advantage in any school or school system.

<sup>1</sup> See Bibliography at the close of the chapter.

FIG. 12.—A SCALE FOR MEASURING ABILITY OF CHILDREN IN GEOGRAPHY IN GRADES 4, 5, 6, 7, AND 8

Grade	G	I	S	U
4	1	4	58	73
5	4	8	73	84
6	6	12	79	88
7	12	21	88	94
8	12	21	88	94

<p>207. Name three agencies or processes at work making rocks into soil.</p> <p>215. By what states would you pass in going by boat from Cincinnati to Memphis?</p> <p>150. Why is most of the rainfall of Australia limited to the eastern and south-eastern coasts?</p> <p>162. Much of India receives from 12 to 16 inches of rainfall in July and less than 1 inch in January. Explain.</p> <p>225. Which is the greater distance and why, 30 degrees west of Washington, or 30 degrees south of Washington?</p> <p>216. New Orleans is in 30 degrees North Latitude and St. Louis is in 39 degrees North Latitude. They are in the same Longitude. About how far apart are they in miles?</p>	<p>115. Name two important valleys of the United States near the Pacific coast.</p> <p>127. Why is mining an important business in the Appalachian region?</p> <p>217. Name five natural wonders of the United States.</p> <p>114. Why is New York so important as a dairying state?</p> <p>132. Why doesn't California grow much corn?</p> <p>172. Why is the Trans-Siberian railroad of so much importance to Russia?</p> <p>173. Why is the Niger river of less importance than the Nile?</p> <p>187. Give one reason why Chicago rather than St. Louis has become the railroad center of the middle west.</p>	<p>52. What is the largest city of your state?</p> <p>64. Where is Alaska and to whom does it belong?</p> <p>84. Name four large cities of Europe.</p> <p>92. Give the capitals of France and Germany.</p> <p>101. Name two large bodies of water that border on Florida.</p> <p>45. Name four things you use for food that do not grow where you live.</p> <p>68. Give one reason why so many of the great cities of the United States are near the sea-coast.</p> <p>72. Which is the coldest and which the warmest part of South America?</p>	<p>63. What country is north of the United States and to whom does it belong?</p> <p>102. Name two other countries in North America besides the United States.</p> <p>24. Name five wild animals.</p> <p>5. What two oceans border on the United States?</p> <p>43. Name a plant used for making cloth.</p> <p>49. Write your whole address.</p> <p>27. Name two kinds of work that men do in getting materials for building houses.</p> <p>26. Name two kinds of work that men do in getting food for us.</p> <p>34. How can you tell from what direction the wind is blowing?</p> <p>50. To whom do the streets or roads belong?</p>
--	---	--	--

**The Boston Tests in Geography.**—These tests were prepared under the guidance of an educator who had given careful consideration to the true aims of geography in the schools. The result is that the two tests, one on the geography of the United States, and the other on the geography of Europe, consist of questions well chosen from the thought standpoint, and questions that are likely to have an influence entirely in the right direction in the teaching of the subject. While the tests have never been fully standardized and are not available, they are of such significance in showing development in the right direction, that it will be worth while to describe them. This can be done in the words of the author.<sup>1</sup>

The test was prepared with a view of ascertaining:

- (a) The character of the geographical knowledge of the pupils tested;
- (b) The ability of the pupils tested to reason from geographical data;
- (c) The relative adequacy of their knowledge of the general geographical features of the United States and Europe; and
- (d) Whether scientific measurement of educational results in geography is possible.

#### THE SCOPE OF THE TEST WHICH WAS GIVEN

It is obvious that a forty-five-minute test can cover only a limited field of geography. Therefore, the test was confined to the most important countries of the world, viz. the United States and the countries in Europe. Although these countries are studied chiefly in the fifth and sixth grades, by no means does it follow that simply fifth and sixth grade work was tested. The study of Europe and Canada in the sixth grade should certainly include the review of many essential features of the geography of the United States. In the seventh grade the work with Asia and Africa should involve not a little review of both the United States and Europe.

<sup>1</sup> See Bibliography at close of the chapter.



Indeed, the makers of a course of study cannot be justified in devoting so much time to Asia and Africa as is the case in our present course, unless such study requires full explanation of the relationship existing between these countries and the more progressive countries of the world. Through the study of such relationship, there is obtained a definite review of many important facts and principles of the geography of the United States and Europe.

#### AIMS OF GEOGRAPHY TEACHING

As is well known, the conception of geography teaching to-day is quite different from that of fifty or even twenty-five years ago. Then the study of the subject consisted largely in memorizing definitions, in learning the location of places, and in learning unrelated facts about the different countries of the world.

At the present time we consider that the value of geography lies not so much in a knowledge of facts concerning the earth and its people as in an understanding of the various ways in which man's activities are influenced by physical environment.

As a result of the study of geography in the elementary school the pupil should gain :

1. An abiding interest in the different peoples of the world, their industries, their achievements, and their relations to ourselves.
2. A mastery of geographic facts and principles sufficient to enable him to explain :
  - (a) The growth of the leading cities of a region.
  - (b) The development of important industries.
  - (c) The dependence of one part of the world upon another.
3. A breadth of mind which will lead to a sympathetic understanding of races and nations other than his own.
4. A working knowledge of the subject by a thorough training in the use of maps, texts, and reference books so that he can work out new problems independently.

In short, geography should help the pupil to interpret his environment, which in the case of civilized man reaches out to all parts of the world.



## QUESTIONS ON UNITED STATES

(An outline map of the United States was printed at the head of the questions.)

1. Locate on the map the cities named at the right :

	<i>Cities</i>	<i>Products</i>
2. In the column marked	Minneapolis . . .	
"products," write opposite the	Pittsburgh . . .	
name of each city the name of	Lowell . . .	
a product for which the city is	New Orleans . . .	
noted.	Duluth . . .	
	Galveston . . .	
	Lynn . . .	

3. Give reasons for the growth of Minneapolis.
4. Below is given a list of articles which we use in our homes. Write below each word the name of the state in which that article is produced in large quantities :

cotton      oranges      cane sugar      rice      coal      iron

5. Write on the map the name of each state which you have just written in answering Question 4.
6. Why do the states just east of the Rocky Mountains receive less rain than Massachusetts?
7. Explain the way in which the flood plains of the Mississippi River have been formed.

## QUESTIONS ON EUROPE

(An outline map of Europe was printed at the head of the questions.)

1. Locate on the map two seaports of European Russia.
2. Why are the seaports of Russia not so important as the seaports of England?
3. Of what value to the countries of Europe are their colonies in other parts of the world?

4. Why does England import large quantities of wheat?
5. Write on the map the names of the leading manufacturing countries of Europe.
6. Why has Germany become very important as a manufacturing country?
7. Why is the climate of Italy different from that of Germany?

The results of the test show that it is possible to ascertain by carefully selected tests whether or not the true aims of geography have been accomplished in the teaching. It is evident that pupils may remember locational facts without being able to use these in any adequate way in answering the questions which occur to one in daily life. This means that locational facts should be properly subordinated to other more vital phases of the subject. The close relationship between questions 1 and 2 in the test on the United States shows the correct method of fixing in mind the location of places through the study of facts which make those places worth remembering. The important consideration is not the locational facts, but the reasons behind them. There is little or no value in knowing the location of places to which no significance is attached.

While it is possible that the standardizing of these tests would in time have worked harmful results in the Boston schools, they do indicate the type of questions that should form the basis of examinations or tests in geography.

**Language Tests.**—Tests devised by Starch<sup>1</sup> and Charters<sup>2</sup> are available for testing the language errors of children. The *Starch tests* are designated "Grammatical Scale A," "Grammatical Scale B," "Grammatical Scale C." The same type of material is contained in each scale, although they are not guaranteed to be of equal value. The use of the three tests, however, would give a more accurate measure of the child's

<sup>1</sup> "Educational Measurements," Chap. 7.

<sup>2</sup> See Bibliography at close of the chapter.

language ability. The tests are built on the plan of choice of words. They are illustrated by the following form (Scale A, Step 5, Sentence 2).

"The Gazette reported (he; him) to be dead."

The pupil is to mark out the incorrect expression, leaving the correct one. The chief criticism of the Starch tests<sup>1</sup> is that they have not been devised on the basis of an extensive study of the errors actually made by children. For instance, in Scale A, there are thirty-seven errors. Eighteen of these do not appear in the Connersville, Boise, and Kansas City studies of the common errors of grade children. Of the other 19, there are 2 on the double negative; 2, sequence of tenses; 3, use of shall and will; and 7, choice between the objective and nominative form of the pronoun. In view of the inadequacy of these tests, it seems unnecessary to go into detail with reference to the scoring or use of the tests. Quite surely, however, the teacher will find in these tests some helpful suggestions. Tentative standards are given by the author, making possible comparison.

The *Charters' Language Test* covers only pronouns. It is based upon a careful study of the pronoun mistakes of school children, as indicated by the collection and study of more than 25,000 errors that pupils made in using pronouns in their oral language. The test consists of sentences, such as:

" 8. Who do you want?

" 16. They made baskets and filled it with holly."

in which there are errors in the use of a pronoun. The pupil is required to write the correct form. The test is designed to be used in grades three to eight, and to measure the pupil's ability to use the correct forms of pronouns. The test is of the right type and is based upon a fundamental study so that there is in it no misplaced emphasis. The teacher will need to use a score sheet for her pupils in order to locate the specific errors made by each pupil.

<sup>1</sup> They should be designated as "language tests," not "grammatical scales."

**Grammar.** — Standard tests in grammar<sup>1</sup> have been devised by Starch and Charters, and one is in preparation by Buckingham. The Starch grammatical tests have for their purpose the measuring of grammatical knowledge — Test 1, parts of speech; Test 2, cases; Test 3, tenses and modes. These tests are of very doubtful value for use in the grades. The studies by Hoyt and Briggs<sup>1</sup> have shown that it is undesirable to teach formal grammar in the grades. The Charters test in grammar is the same as the language test, except that the pupil is required to give the grammatical rules upon which the corrections are based. It is doubtful if the giving of such rules is a desirable practice in the grades. It is not recommended, therefore, that the grammatical section of the Charters test be used.

**Music.** — Seashore<sup>1</sup> has devised a prognostic test of musical talent, which has been perfected, and which should prove of unusual value. For a number of years Dr. Seashore has been refining this test. The test discovers musical sensitivity, musical memory and imagination, musical intellect, and musical feeling, with an accuracy which justifies a definite conclusion with reference to the musical talent of the pupil tested. As knowledge of this test becomes available, there is no doubt that it will be used more and more widely in discovering for further education the students of unusual ability, as well as those who have so little ability that it is useless for them to put further time on the study of music.

#### BIBLIOGRAPHY

1. Rugg, Earle Underwood, "Character and Value of Standardized Tests in History," *School Review*, 27: 757-771, December, 1919. An unusually helpful critical evaluation of present tests in history.
2. Bell, J. Carleton, and McCullum, D. F., "A Study of the Attainment of Pupils in United States History," *Journal of Educational Psychology*, 8: 257-274, May, 1917.

<sup>1</sup> See Bibliography at close of the chapter.

3. Kelly, Truman L., "Educational Guidance. An Experimental Study in the Analysis and Prediction of Ability of High School Pupils," Teachers College, Columbia University, Contributions to Education, No. 71, p. 33.
4. Buckingham, B. R., "Correlation between Ability to Think and Ability to Remember, with Special Reference to United States History," *School and Society* 5: 443-449, April 14, 1917.
5. "Van Wagenen's Scales in United States History." Information, thought, and character judgment scales. Bureau of Publications, Teachers College, Columbia University, New York City. A. S. Barr of 19 South La Salle Street, Chicago, is working towards a series of tests in United States History similar to the Van Wagenen tests.
6. "Starch's American History Test, Series A." Address, University of Wisconsin, Madison, Wisconsin.
7. "Davis' Tests in United States History — Colonial History." Address, S. B. Davis, University of Pittsburgh, Pittsburgh, Pennsylvania.
8. "Raynor's American History Tests." Address, Bureau of Educational Research, University of Illinois, Urbana, Illinois.
9. "Harlan Test of Information in American History." Address, Bureau of Coöperative Research, University of Minnesota, Minneapolis, Minnesota.
10. Ballou, Frank W., "Geography; A Report on a Preliminary Attempt to Measure Some Educational Results," Boston School Document No. 14, 1915. Tests not available.
11. Branom, M. E., and Reavis, W. C., "Completion Test for Measurement of Minimal Geographic Knowledge of Elementary School Children," Seventeenth Yearbook of the National Society for the Study of Education, Part I, pp. 27-39.
12. "Hahn-Lackey Geography Scale." Address, H. H. Hahn, Wayne State Normal School, Wayne, Nebraska. Price, \$.07 per copy.
13. "Starch Geography Test." Price, \$.02 per sheet. Address, The University Supply Association, Madison, Wisconsin.
14. "Charters Diagnostic Test in Language and Grammar for Pronouns." Address, Bureau of Educational Research, University of Illinois, Urbana, Illinois. Price and postage: Language Edition, \$.60 per hundred copies; Grammar Edition, \$.90 per hundred copies.
15. Hoyt, Franklin S., "Studies in English Grammar," *Teachers College Record*, November, 1906.

16. Briggs, Thomas H., "Formal English Grammar as a Discipline," *Teachers College Record*, September, 1913.
17. Starch, Daniel, "Educational Measurements." Chapter 7 on English Grammar. The Macmillan Co.
18. Seashore, C. E., "The Psychology of Musical Talent," Silver, Burdett & Co., Boston.
19. Seashore, C. E., "Vocational Guidance in Music," University of Iowa, 1916, First Series, No. 2.
20. Seashore, C. E., "Avocational Guidance in Music," *Journal of Applied Psychology*, 1: 342-348, 1917.
21. Seashore, C. E., "The Measurement of a Singer," *Science*, 35: 201-212, 1912.



## CHAPTER IX

### THE MEASUREMENT OF HIGH SCHOOL SUBJECTS

It is possible that some day high school pupils will be so distributed, on the basis of ability, and so managed, by teachers who are thoroughly competent and sympathetic, that formal examinations will disappear entirely from the high school, or will appear no longer as formal examinations, but merely as review recitations, thus forming a part of good teaching rather than a part of an examination scheme. As long, however, as more or less formal examinations continue to be given, teachers and educators must continue in efforts to reduce the evils of the examination system by educating teachers in the matter of giving and grading examinations and by seeking in every way possible to standardize the tests which of necessity must be given. The standardized test, in almost every case, is more reasonable, shows fewer idiosyncrasies, and shows more uniformity in grading, than do the ordinary unstandardized examinations. Apparently, therefore, in secondary work, two movements must continue to develop — one looking toward the gradual elimination of the formal examination, the other looking forward to the standardization of such tests as are given.

**High School Tests.** — Can the more strictly high school subjects, such as Latin, French, algebra, geometry, physics, ancient history, chemistry, general science, literature, composition, the commercial subjects, agriculture, home economics, and manual training, be measured by standard tests and scales? The answer is, "Yes, in so far as they are tool subjects or mechanical subjects." Tests have been

developed and more or less fully standardized in algebra, geometry, physics, Latin, German, French, ancient history, and commercial subjects. It is true that, with the development of the junior high school, the line between high school and grade subjects is not so clearly defined as it once was, so that in the junior high school, and even in the senior high school for that matter, some grade tests are being used in diagnosing the condition of the child with reference to the fundamental subjects or determining his efficiency for certain lines of work, particularly the commercial work. For such purposes, therefore, grade tests, such as diagnostic tests in arithmetic, reading for speed and accuracy, writing, spelling, and composition, are used to a considerable extent.

Assuming that the new entrance age for secondary work is twelve or the arrival of adolescence, then we may expect that these tests will be used more and more for determining the condition of the child with reference to the mastery of the fundamentals of grade work.

**Why Tests Have Developed More Slowly in High School Work.** — The work of developing standard tests in the secondary subjects has proceeded more slowly than in the elementary subjects. Since tests are developed chiefly by educational experts in colleges and universities, and since such experts are as interested in the secondary schools as in the elementary schools, in fact more interested from the standpoint of training teachers, it appears that the reason why standard tests and scales have not been developed for high school subjects so rapidly as in the elementary subjects must be due to some intrinsic values or limitations involved in the subjects themselves.

The reasons are: First, most of the high school subjects are not tool subjects. They are of value chiefly because of content and appreciative values. These values are more intangible, more difficult to measure, than the simple elements

involved in the tool subjects. Second, the old academic viewpoint that secondary work is merely preparatory is changing. The old viewpoint made the mastery of subject matter, as such, the essential consideration. The present tendency, however, is to minimize the importance of high school work as merely preparatory, to look more towards use and application, and to make of the high school a real people's school serving the broader aims of education. The efficiency of work on this basis cannot be tested nearly so well by examination methods. Even a subject like mathematics does not become a tool subject for a large percentage of pupils. Apparently, therefore, appreciative values and an understanding of the subject from the standpoint of enjoyment and perspective are just as important as the mere mastery of subject matter. Third, it is in subjects like literature and history especially that the fact, subject matter basis is particularly undesirable. Literature, to be effective and to carry over into later life, must be taught on a basis of appreciation and enjoyment. It does not lend itself to rigid testing. History, likewise, deals with life problems, which depend for their development upon present-day problems, pupil interests, community contacts, and teaching equipment: So that any attempt to reduce history to a mere mechanical basis renders it of little value.

In short, standard tests and scales have proven of value chiefly in measuring the tool subjects and the mastery of subject matter. The high school curriculum has many other values, some of which are possibly even more important than the strictly measurable ones. It will be worth while, however, to note the development of scales in high school subjects in so far as they have developed.

**Algebra.** — It is generally considered that algebra in its more fundamental processes is a tool subject which may be measured with a degree of accuracy approximating the simpler fundamental processes in arithmetic. The attempts to

formulate scales or standard tests in algebra have proceeded with greater rapidity than any other high school subject. Among the tests available are: the Monroe "Standardized Research Tests in Algebra," the Rugg and Clark "Standardized Tests in First Year Algebra," and the Hotz "First Year Algebra Scales."

1. *The Monroe Tests* are based upon the assumption that the equation is the central fact of algebra. The tests cover the simple operations as follows: Test 1, removal of parentheses; test 2, clearing the equation of fractions; test 3, solving for  $x$ , a special case of division; test 4, transposition; test 5, collecting terms; test 6, solution of simple equations.

2. *The Rugg and Clark Standardized Tests in First Year Algebra* have been very generally used. They have been quite fully standardized. They will be found of interest to teachers of algebra, and, if properly used, will be of considerable value.

The chief criticism of the tests is that they attempt to cover the subject as fully as it is covered by textbooks and, on the whole, seem a little more difficult than necessary. Even so, they will be found much simpler and much more reasonable than tests which teachers ordinarily give. This will be apparent from an examination of a few of the tests.

Test 4, in simple equations, follows herewith.

1. $2x \div 3 = 11$	Answer.....
2. $4c = 6c + 12$	Answer.....
3. $5x - 3 = -20$	Answer.....
4. $13 = 2x - 8$	Answer.....
5. $12x - 7 - 15x = 10$	Answer.....
6. $3x + 4 = 16$	Answer.....
7. $8p = 15p + 14$	Answer.....
8. $7x - 6 = -29$	Answer.....
9. $17 = 3x - 8$	Answer.....
10. $9x - 8 - 13x = 7$	Answer.....
11. $5x + 2 = 27$	Answer.....
12. $6t = 9t + 21$	Answer.....

13. $8x - 7 = -35$	Answer.....
14. $19 = 5x - 9$	Answer.....
15. $21x - 12 - 26x = 14$	Answer.....
16. $4x + 5 = 17$	Answer.....
17. $11s = 13s + 20$	Answer.....
18. $9x - 6 = -40$	Answer.....
19. $23 = 6x - 9$	Answer.....
20. $15x - 11 - 21x = 18$	Answer.....
21. $6x + 3 = 33$	Answer.....
22. $12y = 15y + 12$	Answer.....
23. $10x - 7 = -33$	Answer.....
24. $21 = 8x - 5$	Answer.....
25. $18x - 3 - 23x = 9$	Answer.....
	Number attempted.....
	Number right.....

Test number 5, parentheses, consists of 42 examples, as follows:

1. $6(3x + 8)$	Answer.....
2. $5(4x - 2)$	Answer.....
3. $-3(4x - 2)$	Answer.....
4. $-4(3x - 4)$	Answer.....
5. $9(-7x - 1)$	Answer.....
6. $-8(-4x - 7)$	Answer.....
7. $8(5x + 4)$	Answer.....
8. $7(4x - 3)$	Answer.....
9. $-5(6x + 7)$	Answer.....
10. $-6(4x - 8)$	Answer.....
11. $7(-8x - 3)$	Answer.....
12. $-9(-3x - 6)$	Answer.....
13. $7(4x + 6)$	Answer.....
14. $6(3x - 5)$	Answer.....
15. $-4(7x + 5)$	Answer.....
16. $-5(6x - 7)$	Answer.....
17. $7(-9x - 2)$	Answer.....
18. $-9(-5x - 4)$	Answer.....
19. $5(6x + 7)$	Answer.....
20. $7(-9x - 2)$	Answer.....
21. $-5(3x + 8)$	Answer.....

22. $-7(8x-5)$	Answer.....
23. $8(-6x-3)$	Answer.....
24. $-6(-5x-4)$	Answer.....
25. $8(7x+5)$	Answer.....
26. $5(3x-6)$	Answer.....
27. $-4(6x+3)$	Answer.....
28. $-5(7x-4)$	Answer.....
29. $9(-6x-2)$	Answer.....
30. $-7(-5x-3)$	Answer.....
31. $4(3x+7)$	Answer.....
32. $6(7x-3)$	Answer.....
33. $-5(2x+8)$	Answer.....
34. $-4(3x-7)$	Answer.....
35. $8(-7x-3)$	Answer.....
36. $-7(-3x-5)$	Answer.....
37. $7(5x+9)$	Answer.....
38. $5(8x-4)$	Answer.....
39. $-6(5x+2)$	Answer.....
40. $-9(4x-6)$	Answer.....
41. $7(-6x-4)$	Answer.....
42. $-8(-4x-6)$	Answer.....
	Number attempted.....
	Number right.....

Test number 8, factoring, contains 25 examples, as follows:

1. $5x^2+15x^3$	Answer.....
2. $a^2-64$	Answer.....
3. $y^2-6y+9$	Answer.....
4. $b^2+11b+28$	Answer.....
5. $5x^2+16x+3$	Answer.....
6. $6a^2+9a^3$	Answer.....
7. $x^2-16$	Answer.....
8. $t^2-16t+64$	Answer.....
9. $a^2+12a+27$	Answer.....
10. $9x^2+36x+32$	Answer.....
11. $9y^2-6y^3$	Answer.....
12. $b^2-25$	Answer.....
13. $a^2-14a+49$	Answer.....



14. $m^2 + 12m + 35$	Answer.....
15. $7x^2 + 26x + 15$	Answer.....
16. $8x^2 - 12x^4$	Answer.....
17. $y^2 - 81$	Answer.....
18. $p^2 - 18p + 81$	Answer.....
19. $c^2 + 11c + 30$	Answer.....
20. $3x^2 + 22x + 35$	Answer.....
21. $12y + 18y^4$	Answer.....
22. $p^2 - 25$	Answer.....
23. $x^2 - 10x + 25$	Answer.....
24. $a^2 + 10a + 21$	Answer.....
25. $5x^2 + 28x + 15$	Answer.....
	Number attempted.....
	Number right.....

The time allowance for test 4, is 3 minutes, for test 5, 2 minutes, and for test 8, 4 minutes. Time limits have been fixed for each test and tentative standards have been determined. For example, the average of 27 schools on test 5 was 10.4 examples attempted with 9.7 correct.

3. *The Hotz First Year Algebra Scales* have been worked out on a scientific basis and the problems located on a point scale. This is an advantage to the extent that teachers understand its value. The problems differ from those in a standard test in that they are of increasing difficulty. The pupil is measured by the point which he reaches on the scale. This requires the scale to grow sufficiently difficult that he will at some point fail to go farther. The test in multiplication and division, series B, which follows herewith, is illustrative.

#### MULTIPLICATION AND DIVISION

Carefully perform the operations as indicated. Reduce all answers to their *simplest* forms.

1.  $3 \cdot 7y =$

2.  $\frac{12n}{4} =$

3.  $2a \cdot 4ab^2 =$
5.  $\frac{2}{3}$  of  $9m =$
7.  $4x \cdot (3xy^3) =$
9.  $\frac{18m^2n - 27mn^2}{9mn} =$
11.  $(2a^2 + 7a - 9)(5a - 1) =$
13.  $\frac{7a}{15} \div \frac{7a^2}{20} =$
15.  $\frac{m+n}{a} \cdot \frac{b}{m^2-n^2} =$
17.  $\frac{c^4-d^4}{(c-d)^2} \cdot \frac{c-d}{c^2+d^2} =$
19.  $\frac{a^2 + \frac{3}{2}a - 1}{a+2} =$
20.  $\frac{p^2+4p-45}{p^2+2p+4} \cdot \frac{p^3-8}{p^2-81} \cdot \frac{1}{3pr-15r} =$
21.  $\frac{x^3+27}{x^2+x-12} \div \frac{3x+9}{x+4} =$
22.  $64\frac{2}{3} \times 27\frac{1}{3} =$
23.  $\frac{3\sqrt{6}a}{2a\sqrt{18}} \cdot 12a =$
4.  $6c^3 \div 2c^2 =$
6.  $\frac{-8a^2b}{4a^2} =$
8.  $a^3 \cdot (-3a) \cdot (-2a) =$
10.  $\frac{4x^4}{5} \div 2x^2 =$
12.  $\frac{n^4+7n^2-30}{n^2-3} =$
14.  $\frac{-12x^2y^2 \cdot (x-2)}{3x^2y^2} =$
16.  $(-3xy^3)^4 =$
18.  $3x^{\frac{1}{2}} \cdot 4x^{\frac{1}{2}} =$

There is no doubt that any of these tests in algebra will prove of value. They have been standardized, they permit comparison, they will be valuable for research purposes, and they have the advantage when used by teachers for promotion purposes of avoiding the unusually difficult problems often used by teachers in final examinations. In other words they are more reasonable than tests usually given by teachers. Teachers frequently have erroneous ideas about the promotion of children; some even think it to their credit to fail a large number of pupils. If a pupil can pass simple tests such as the Monroe tests, he should be

permitted to go forward with advanced work. While the Monroe tests cover only the simple fundamental processes of first year algebra, the Rugg and Clark tests cover the entire field of secondary algebra. The tests will distribute pupils so as to show a teacher that she is instructing a group of pupils who differ widely in ability and need help and drill on widely varying details. The wise teacher of algebra will keep for every pupil a card showing his mistakes or weaknesses, such as mistakes in sign, errors in copying, errors in factoring, etc. The standard tests will further aid in locating pupils' weaknesses. Every teacher of mathematics in a high school and every superintendent should become familiar with at least one of the available tests in algebra.

**Geometry.**—The *Stockard and Bell*<sup>1</sup> test in geometry consists of 70 questions arranged in 20 groups. In devising the test "the attempt was made to call for information that is to be found in all standard textbooks; to test for important and fundamental principles of geometry; to provide such a range of questions as to be representative of the whole field of elementary geometry, and to include memory facts, knowledge of content, organization of subject matter, and particularly ability to do originals; and to confine the list to such dimensions that every question could be tried by the average high school pupil in 40 minutes."

The 20 groups "involve drawing figures, naming figures, indicating order of development in demonstrations, completing statements, stating the converse, definitions, regular polygons, parts of a demonstration, angular relations, area of a trapezoid, angles in polygons, angles in circles, congruency, similarity of triangles, loci, auxiliary lines, simple constructions, ratio and proportion, algebraic expression of geometric relations, and equivalent construction."

It is evident that the authors have attempted to measure quite fully the student's mastery of the subject matter of

<sup>1</sup> See Bibliography at close of chapter.

elementary geometry. The test was given to 372 school students who had completed a year's work in geometry. About one third of the pupils tested were able to attempt all of the questions. On the basis of the tests given, the different questions are rated. The authors think the test not very practical for general high school use. It is too lengthy and, on the whole, a little too difficult. Teachers may use it, however, for diagnostic purposes or purposes of research.

**Diagnostic Tests in Mathematics.**—Six tests have been selected by Anna L. Rogers.<sup>1</sup> The tests, together with the time for explaining and giving the same, are as follows:

(1) Algebraic Computation . . . . .	12 minutes
(2) Interpolation . . . . .	18 minutes
(3) Geometry . . . . .	40 minutes
(4) Superposition . . . . .	7 minutes
(5) Mixed Relations . . . . .	8 minutes
(6) Trabue Scales, L and J . . . . .	12 minutes
Total . . . . .	97 minutes

The total time required, 97 minutes, is just a little more than 2 regular high school periods. Yet in that brief time these tests enable a competent teacher to diagnose the mathematical ability of ninth grade pupils with a view to improvement in the classification of students in the high school by eliminating from the mathematics classes those unfit for further mathematical training, and selecting those capable of progressing at a more rapid rate than the majority. The tests also serve to discover particular lines of mathematical weakness. They may be given in the seventh and eighth grades, but, when so given, the time limits must be considerably extended. These tests are "designed to measure the more important phases of mathematical capacity demanded by high school mathematics, and, in particular, the ability to manipulate numerical and algebraic symbols, the ability to

<sup>1</sup> See Bibliography at close of chapter.

grasp and handle spacial relations, and the ability to deal effectively with words. They are of such a nature as to enable an intelligent teacher to form an independent estimate of the pupil's mathematical capacity and likelihood of success in further lines of mathematical work. They measure original ability rather than effect of training."

Miss Rogers has given directions for giving the tests and evaluating the scores, and has fixed tentative standards. She says, "As tentative standards, we suggest: (1) Where a pupil's score is greater than 150, he has capacity to progress at a more rapid rate than the ordinary high school student. (2) Where a pupil's score is less than -150, he shows incapacity to progress in mathematics at the rate of the ordinary high school student and, other things being equal, should be released from further training in the subject." The group coming between -150 and +150 is considered the normal group in high school mathematics.

**Henmon's Latin Tests.** — Prof. V. A. C. Henmon of the University of Wisconsin has developed a series of vocabulary tests, A, B, C, and D, which are of equal difficulty and in which the words are arranged in order of difficulty. All of the words in these tests have been carefully evaluated. A series of sentence tests is also available, consisting of tests 1 and 2, of equal difficulty, and test 3, in which the sentences are all of approximately the same difficulty. Standard scores are given for both the vocabulary and the sentence tests. One test is required for each pupil.

The Henmon tests in Latin have several advantages over ordinary tests given by teachers. In the first place, they are scientifically constructed, and they are based upon vocabulary which is common to Caesar, Cicero, Virgil, and 13 of the most frequently used first year Latin texts. In the second place, the tests are thoroughly standardized, making possible accurate grading and comparison with other schools, with other classes, or among pupils of the same class. In the third

place, such tests are helpful in the study and analysis of class work. This is well illustrated in the overlapping of abilities in successive years as shown by these tests. In test A the medians for the first year range, for different classes, from 1 to 11, the median for all first year classes being 4. In the second year the class medians range from 4 to 19, the median for the year being 7. In the third year the class medians range from 3 to 25, the median for the year being 20. If corresponding medians for various schools show such wide variation, the individual scores must evidently show very much greater range. The overlapping of abilities in Latin is thus seen to be comparable to the overlapping of abilities in other subjects.

The administering of standard tests of this kind should discover the pupils of exceptional interest and ability on the one hand, and pupils on the other hand who, through lack of ability or interests, do so poorly that it is useless to have them continue the study.

In so far as Latin is a tool subject, standard tests are applicable. In so far as the subject is of interest chiefly because of other values, to that extent the teacher should be cautious in using standard tests or should use them only for her own enlightenment, being careful that they do not formalize her work. This will mean that the results of the tests are in general not brought to the attention of the pupils.

**Physics.** — A physics test has been devised by Professor Daniel Starch, University of Wisconsin. It consists of 75 mutilated sentences. They cover the 102 facts, principles, and laws of physics which the author has determined upon as the most essential. The basis for the determination was an examination of 5 widely used textbooks. The tests are easily administered. The value of the tests, however, has not been demonstrated. It is doubtful if physics can be reduced to facts in such formal fashion as this test would suggest.



**Commercial Tests.** — In the volume, "Commercial Tests and How to Use Them," Sherwin Cody has brought together a summary of the use of tests in determining the relative standing of students graduating from the commercial departments of high schools. Tests are available, covering the following subjects :

- (1) Tabulating, mental alertness.
- (2) Reproducing instructions — designed to test memory and natural industry.
- (3) Invoicing.
- (4) Fundamentals of arithmetic, an adaptation of the Courtis tests.
- (5) Business arithmetic, including fractions, trade extensions, and percentage.
- (6) English, including spelling, elementary language, advanced language, elementary punctuation, and advanced punctuation.
- (7) Letter writing.
- (8) Answering letters.
- (9) Stenographic tests covering transcribing and typewriter copying.
- (10) Copying for the mimeograph.
- (11) Addressing envelopes with the pen and filing.

For each of the above lines of testing there are duplicate tests, full directions for administering the tests, keys for grading, and tentative standards. The need of such commercial tests is evident to those who have attempted to select stenographic help or to evaluate the products of various commercial schools. The test in transcribing, for instance, is quite definite. The student is expected to transcribe a standard business letter of 300 words in 5 minutes. This means 60 words per minute. In many high schools, where there is no particular standard, students are permitted to graduate with a speed in transcribing of only 30 words per minute or even less. In like manner, the student can be accurately

checked on ability to file, answer a letter, spell, or use the fundamental processes in arithmetic. One of the main values of Mr. Cody's work is to suggest standards for commercial work, and all admit that this is a type of work which can be standardized.

**Sackett's Ancient History Scale.**—Some will doubt the value of this test because it attempts to reduce a thought subject to a mechanical basis. Professor Sackett has at least shown the difficulty of formulating an ancient history scale. His work is handled on an approved scientific basis, but if the premises are faulty the conclusion, of course, cannot be other than erroneous. What we want our students to get from ancient history is not a memory mastery of the facts, but, instead, an appreciation of the problems and the development of a method for the solution, not only of the problems of ancient history, but of present-day history.

The scale (so called) consists of eight tests, each containing ten points. Test No. 1, which is typical, is as follows:

For what are the following men noted?

- (1) Hannibal.....
- (2) Khufu or Cheops.....
- (3) Demosthenes.....
- (4) Darius.....
- (5) Solon.....
- (6) Charlemagne.....
- (7) Attila.....
- (8) Constantine.....
- (9) Mithridates.....
- (10) Justinian.....

It is evident that this is a fact testing scale. Its general acceptance would quite surely be detrimental to the proper teaching of history. A student might have all knowledge as tested by this scale, and still be entirely lacking in the spirit and method of history. The test should be used with

11. Chapman, J. Crosby, "The Measurement of Physics Information," *School Review*, 27: 748-756. (This is a fact or information test.)
12. Uniform Science Tests in Physics," by Franklin T. Jones, University School, Cleveland, Ohio.  
*Reference:* *School Review*, 26: 341-348, May, 1918. The subjects covered are: thermometers, fusion, vaporization, specific heat, heat exchange.
13. Chemistry Scales.  
*References:* Bell, J. Carleton, "A Test in First-Year Chemistry," *Journal of Educational Psychology*, 9: 199-209, April, 1918. Webb, Hanor A., "A Preliminary Test in Chemistry," *Journal of Educational Psychology*, 10: 36-43, January, 1919.
14. Cody, Sherwin, "Commercial Tests and How to Use Them," World Book Company, Yonkers-on-Hudson, New York.
15. Sackett, L. W., "A Scale in Ancient History," *Journal of Educational Psychology*, 8: 284-293, May, 1917.
16. Rugg, H. O., "A Scale for Measuring Freehand Lettering for Use in the Secondary Schools and Colleges." Address, H. O. Rugg, School of Education, University of Chicago, Chicago, Illinois. Price, \$.25 a copy.  
*Reference:* Rugg, H. O., "A Scale for Measuring Freehand Lettering," *Journal of Educational Psychology*, 6: 25-42, January, 1915.
17. "Tests in Home Economics," Supplementary Educational Monographs, No. 6 of Vol. 2, University of Chicago Press.
18. Murdock, Katharine, "The Measurement of Certain Elements of Hand Sewing." (Sewing scale is included.) Teachers College, Columbia University, Contributions to Education, No. 103.
19. "A Brief Bibliography of Tests in High School Subjects," *School Review*, 27: 799-809, December, 1919. Covers tests in Latin, Mathematics, Science, and Home Economics.
20. "Kansas Silent Reading Test No. 3." Ability of high school students to read silently. Address, Bureau of Efficiency and Measurement, Emporia, Kansas.
21. "Standardized Tests in Silent Reading No. 3." For high school students. Address, Bureau of Efficiency and Measurement, Emporia, Kansas.

## CHAPTER X

### THE MEASUREMENT OF GENERAL INTELLIGENCE

"THIS is an exceptionally slow class," and "I have such a large number of children this year who are very slow," are expressions which the supervisor frequently hears as he passes from one classroom to another. These statements are based on actual facts in that they describe the condition of many children who are not making progress. The problem is a very real one to the teacher.

It is also not infrequent to hear a teacher say, "I have an exceptionally bright class this year," or "I have 4 or 5 children who are far ahead of the others." It is unusual, however, to hear such statements followed with the remark, "I think certain children in this grade ought to be advanced to another grade." The latter situation should become more common.

In either case the problem is one of knowing the child's mental age or general intelligence. He has been assigned to a certain grade chiefly on the basis of his chronological age or the number of years in school. He has been asked to do a certain type of work because children of his age who are normal are supposed to be able to do it. It may be that such children are being held back when they should be advanced, or the subject matter which is being presented in the regular class is not suited to them, and they are, therefore, in need of a different kind of subject matter presented in a special class.

Sufficient information is available to show that general intelligence tests can be used to determine the mental ages of

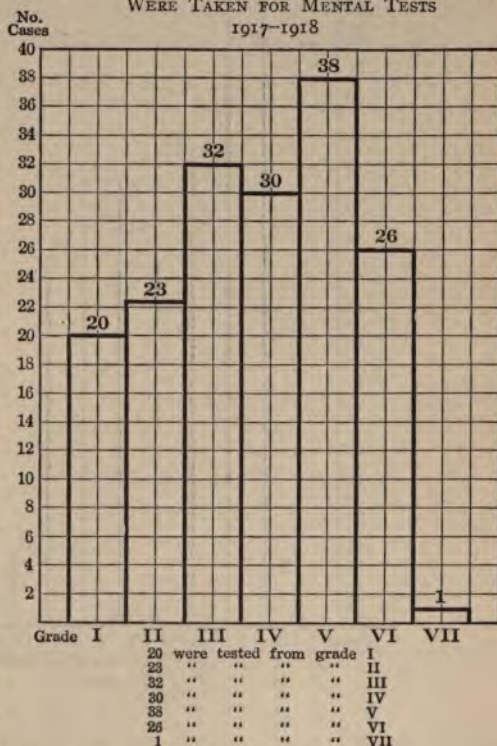
children, and that this information can be used as a basis for reclassification of the pupils. In a city school system in which the Binet-Simon (old form) Test was being used, it was found that a great many children who were too old chronologically were also too old mentally for their grade, *e.g.* when children enter the first grade at the age of six, they are according to correct practice of normal age chronologically if they are 6 or 7 years old in the first grade, 7 and 8 years old in the second grade, etc. If a child in the first grade is 8 years old, he is 1 year too old chronologically. If he tests 8 years old mentally he is also 1 year too old mentally. He should, therefore, be in the second or third grade, provided his school life has been normal. If not, he should receive the special attention that will place him with his appropriate group as soon as possible. In the school system referred to above, out of less than 500 children tested in 4 schools, 60 such children were found. All of these children were advanced one grade and all were able to continue with a good class standing in the grade to which they were promoted.

In a small Iowa city 39 children out of a total of 177 children in 4 schoolrooms were advanced in the fall of 1915, on a basis of a good class standing and being too old for the grade in which they were working. These children were selected from grades 5, 6, 7, and 8. Eleven of them were advanced one half of a grade, 28 were advanced a full grade. At the end of the first semester, February, 1916, not a single pupil failed in gaining promotion, and only 3 received a class standing of less than 85%. The fact that such a large number of children in a school system can be advanced to advantage on a very general measurement, and that so little material is available showing the results of such practice, is indicative that a general intelligence test which will measure in very definite terms the general intelligence of large groups of children is greatly needed, and can be used to prevent waste and unsatisfactory results in classroom practice.



General intelligence tests have also been used for a number of years to select children who are subnormal to be placed in special classes in which they will receive a different type of instruction. Before such tests were available these children were compelled to try to do the same work as the normal child, which resulted in failure. After numerous repetitions had occurred it was realized that they could not advance. They were, therefore, assigned to some other class in which they were permitted to do the thing they were able to do, but only after the loss of much time and energy to both pupil and teacher. With the aid of a general intelligence test these children can be located early in their school classes, thus reducing greatly the amount of waste.

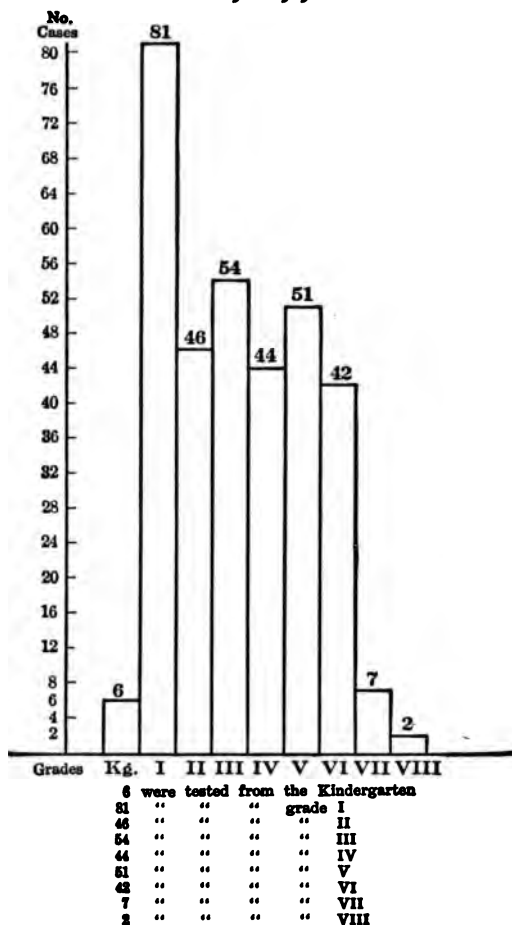
FIG. 13.—GRADES FROM WHICH 170 CHILDREN WERE TAKEN FOR MENTAL TESTS 1917-1918



This tendency is clearly illustrated from Figures 13 and 14, which show the grades from which the children were selected to supply the special classes for subnormal children in a city of over 100,000 after these classes had been in operation two years.



FIG. 14.—GRADES FROM WHICH 333 CHILDREN WERE TAKEN FOR MENTAL TESTS 1918-1919

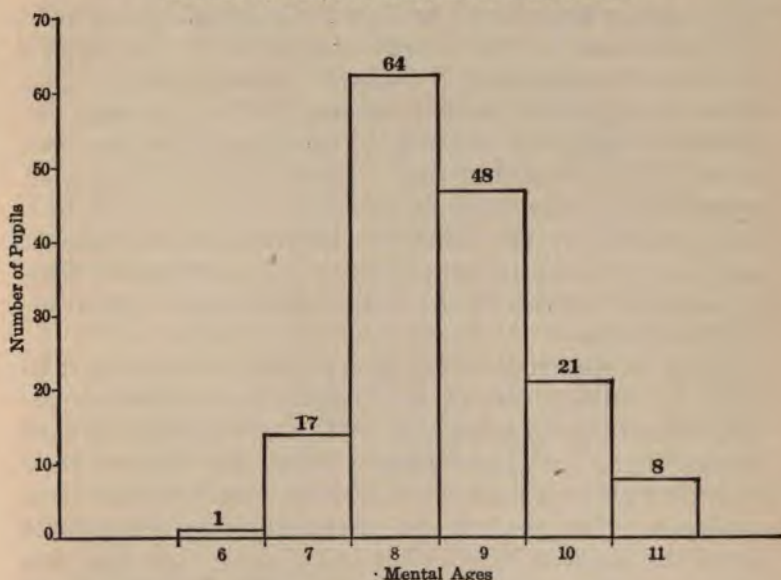


The error which has brought about the necessity for re-classification in the regular classes or in the special classes has resulted from an apparent assumption that children of a certain age group are of the same mental age. This principle

is seen in the practice of entering 6-year-old children into the first grade. At the end of one school year or at the age of 7 years they are supposed to be ready for the second grade, and so forth.

That all children of the same age are not of the same mental age is apparent from the following graph showing the mental ages of 159 unselected children, all of chronological age 9.

FIG. 15. — DISTRIBUTION ACCORDING TO THEIR MENTAL AGES OF 159 UNSELECTED CHILDREN (9 YEARS OF AGE), BINET-SIMON SCALES.



These 159 children make up the entire group of children 9 years of age in a group of 743 children in 4 elementary schools who were tested with the Binet-Simon Test. Instead of all having a mental age of 9 years, only 48 showed exactly this age, 21 children had a mental age of a normal child at 10 years of age, and 8 children a mental age of a normal child at 11 years of age, while 64 children showed a mental

age of a normal child 8 years, 17 children a mental age of 7 years, and 1 child a mental age of 6 years of age. These figures show the same differences in abilities of children as are found in all unselected groups whenever a test that will distribute mental age is used.

Although much has been written about the wide differences in the abilities of children and sufficient evidence given to support such conclusions, yet classroom practice too often lags behind because of the slowness with which usable tests have been developed and placed in the hands of the teachers. The teacher finds it exceedingly difficult to see and to plan for the needs of the class as a whole. Likewise, the principal in his organization of classes is too much given to an organization on a basis of the needs of the group instead of the needs of the individuals in the group. Splendid attempts and good progress have been made in many places to recognize in a very definite way the individual difference among children, but in general, educational practice continues to handle them en masse on account of the lack of knowledge of group intelligence tests.

A test in the hands of the teacher that will enable her to know the mental ages of the children in her class as she knows their chronological ages, will be a long stride toward classroom practice which handles children on a basis of their individual needs instead of group needs. Such tests are now available. The teacher can determine the mental ages of all the children in her class in the same time that was formerly required by an expert to determine the mental age of a single child.

The following group intelligence tests are now available: Trabue Language Scales, not devised originally to measure general intelligence, but found to do so with an accuracy that makes them very valuable; the Otis Group Intelligence Scale; Haggerty's Intelligence Examinations, Delta 1 and Delta 2; and Whipple's Group Tests for the Grammar Grades.

TRABUE LANGUAGE SCALES

**Aim.** — It has been found that the Trabue Language Scales will measure the general ability of pupils.

**Description of Tests.** — The Trabue Language Scales consist of scales B, C, D, E, J, K, L, and M. Scales B, C, D, and E are practically equal to one another. They are intended to be used in pairs — B with C, and D with E — by teachers in determining the abilities of children between the ages of 7 and 20. Each scale is made up of 8 to 10 sentences with words omitted which are to be supplied. "The first sentence in each of these 4 scales is about 1 unit above an arbitrary zero point, the second sentence is approximately 1 unit more difficult than the first, and so on until the last sentence in each scale is about 11 units above zero."

Scales J and K are intended to measure the ability of adults, and are, therefore, of practically no value for public school purposes.

Scales L and M, which are not equivalent to scales J and K, or to any other pair, are intended to measure the abilities of high school students. "They have no very easy sentences, and the differences between the sentences are relatively small."

The grouping into pairs of these scales of equal value provides a duplicate test for checking.

Below is given a copy of Language Scale B to show the nature of these scales:

Name.....  
Write only one word on each blank. Grade.....  
Time limit, seven minutes. Age (on last birthday).....

TRABUE

LANGUAGE SCALE B

1. We like good boys — girls.
2. The — is barking at the cat.
3. The stars and the — will shine tonight.
4. Time — often more valuable — money.

5. The poor baby — as if it were — sick.
6. She — if she will.
7. Brothers and sisters — always — to help — other and should — quarrel.
8. — weather usually — a good effect — one's spirits.
9. It is very annoying to — — tooth-ache, — often comes at the most — time imaginable.
10. To — friends is always — the — it takes.

**Giving the Test.** — The process of giving the Trabue Language Scales is very simple. A preliminary test is provided with simple sentences in order to make clear to the children exactly what they are expected to do. After this preliminary test has been given and explained fully, the children are ready for the regular test.

Each child is provided with a scale on which he writes his name, grade, and age. On each sheet is the instruction, "Write only one word on each blank." Attention should be called to the fact that he will be given a time limit (7 minutes for scales B, C, D, and E; 5 minutes for L and M) in which to do as much as he can. He will possibly not be able to fill all the blanks. As soon as the time limit has expired see that every child stops work and all papers are collected.

**Scoring Results.** — For the convenience of the teacher and the accuracy of the results, the author has provided a detailed scheme for scoring the answers to the different sentences, which is as follows:

#### GENERAL SCHEME

##### Score 2

"A score of 2 points is to be given each sentence completed perfectly. Errors in spelling, capitalization, and punctuation should not be allowed to affect the score.

##### Score 1

"A score of 1 is to be given each sentence completed with only a slight imperfection. A poorly chosen word or a common gram-

mathematical error, which makes the sentence less than perfect and yet leaves it with reasonably good sense, should serve to reduce the score from 2 to 1.

Score 0

"A score of 0 is to be given if the sentence as completed has its sense or construction badly distorted. A sentence must have reasonably good meaning and express a sentiment which might honestly be held by an intelligent person in order to receive a higher credit than zero."

The following is a sample of the answers provided in this scheme, for the first three sentences of Scale B:

LANGUAGE SCALE B

1. We like good boys — girls.  
Score 2  
and, an  
Score 1  
or, not, and good, also.  
Score 0  
for, with, said the, and the.
2. The — is barking at the cat.  
Score 2  
dog, hound, pup.  
Score 1  
dogs, boy.  
Score 0  
man, cat, god.
3. The stars and the — will shine to-night.  
Score 2  
moon.  
Score 1  
light, planets, lights.  
Score 0  
dipper, stripes, clouds, city, sky, sun.  
etc.

The score for each sentence should be recorded on the margin of the test. After all the test papers have been scored, the



scores are transcribed to a Class Record Sheet. Below is given a copy of the Class Record Sheet provided by the author with the scores from a 4-A class in a city school system.

TABLE 49  
CLASS RECORD

Date Jan. 14 '20. Completion Test—Language Scales. Language Scale B. City D. State M. School F. Room No. 9. Grade 4-A. Teacher H. S. Test Given by H. S. Number of Pupils taking test 35. Number regularly enrolled pupils not taking this test 0.

Test began at 11:30, closed at 11:37. Time allowed 7 min. Unusual conditions which might influence results of this test, none. Scores Assigned by H. S. recorded by H. S.

NAMES OF BOYS	AGE		SCORE ON EACH OF SENTENCES										TOTAL SCORE	NAMES OF GIRLS	AGE		SCORE ON EACH OF SENTENCES										TOTAL SCORE
	Yr.	Mo.														Yr.	Mo.										
S. L. . .	10	2	0	2	0	0	0	0	0	0	0	0	2	L. G. . .	9	9	2	2	2	2	0	0	0	0	0	0	8
W. B. . .	11	3	0	2	2	0	0	0	2	0	0	0	6	G. M. . .	10	6	0	2	2	2	2	0	0	0	0	0	8
A. A. . .	11	0	2	2	2	0	0	0	0	0	0	0	6	M. D. . .	11	5	2	2	2	0	0	0	2	0	0	0	8
M. L. . .	10	8	2	2	2	0	0	0	0	0	0	0	6	E. B. . .	12	0	2	2	2	0	0	0	0	0	0	0	8
W. K. . .	9	9	2	2	2	2	0	0	0	0	0	0	8	S. H. . .	10	4	2	2	2	0	0	0	0	0	0	0	8
H. S. . .	11	0	2	2	2	0	2	0	0	0	0	0	8	C. M. . .	9	8	2	2	2	2	0	1	0	0	0	0	9
R. D. C. .	13	0	2	2	2	2	2	0	0	0	0	0	10	L. B. . .	11	10	2	2	2	2	0	0	0	0	0	0	10
H. S. . .	12	8	2	2	2	2	2	0	0	0	0	0	10	E. H. . .	9	3	2	2	2	2	0	0	0	0	0	0	10
A. B. . .	12	6	2	2	2	2	2	0	0	0	0	0	10	E. Ha. .	10	9	2	2	2	2	0	0	0	0	0	0	10
W. Q. . .	10	8	2	2	2	2	2	0	0	0	0	0	10	L. J. . .	9	8	2	2	2	2	0	0	0	0	0	0	10
F. R. . .	10	0	2	2	2	2	2	1	0	0	0	0	11	P. D. . .	11	3	2	2	2	2	0	2	0	0	0	0	10
R. S. . .	10	3	2	2	2	2	2	0	0	0	0	0	12	C. R. . .	10	0	2	2	2	2	0	1	0	0	0	0	11
S. L. . .	10	4	2	2	2	2	0	1	1	2	0	0	12	R. A. . .	10	5	2	2	2	2	1	2	0	0	0	0	11
J. F. . .	12	9	2	2	2	2	2	2	0	0	0	0	12	R. T. . .	10	1	2	2	2	0	2	1	2	0	0	0	11
M. B. . .	9	7	2	2	2	2	2	2	0	0	0	0	12	S. C. . .	9	1	2	2	2	2	0	2	0	0	0	0	12
W. A. . .	14	4	2	2	2	2	2	1	0	0	0	0	13	E. L. . .	10	3	2	2	2	2	1	2	1	0	0	0	12
A. G. . .	10	0	2	2	2	2	2	2	1	0	0	0	15	G. P. . .	10	4	2	2	2	2	1	2	2	1	1	0	15
														L. J. . .	10	8	2	2	2	2	2	2	2	0	0	16	

This record is read as follows: S. L., age 10 years, 2 months, made a score of 2 on the second sentence and a score of zero on all the others which gives him a score of 2; W. B., age 11 years 3 months, made a score of 2 on sentences 2, 3, and 6 and a score of zero on all the others making a total score of 6; etc. These scores are then distributed and the class median is determined, which for this class is 10.6.

**Interpreting and Using Results.** — After the class records have been made and the class scores determined, the next problem for the teacher is to interpret her results and apply them to her classroom practice. This can be explained best by reference to concrete situations.

The first point to be determined is the relation of the class score to any class standards. The class standards for the Traubue Language Scale as reported by the author are as follows:

TABLE 50  
STANDARD LANGUAGE SCALE SCORES

SCALES B, C, D, E, F.	GRADE OR CLASS	SCORE (Median)	HALF	SCORE
	II	4.8	IIB IIA	3.8 5.8
	III	8.0	IIIB IIIA	7.4 8.6
	IV	10.0	IVB IVA	9.6 10.4
	V	11.4	VB VA	11.1 11.6
	VI	12.4	VIB VIA	12.1 12.6
	VII	13.4	VIIB VIIA	13.1 13.6
	VIII	14.4	VIIIB VIIIA	14.1 14.6
			Tentative Standards in Scales J & K                      L & M	
	H. S. I	15.2	7.5	7.5
	H. S. II	16.0	8.6	9.2
	H. S. III	16.7	9.4	10.5
	H. S. IV	17.4	10.0	11.5

The class score for the 4-A grade reported on the above class record sheet, Table 49, is 10.6. The standard score for this grade is 10.4. This class is, therefore, slightly above the standard.

When the test has been given in other classes of the same city further comparisons can be made. Below are given the class scores from the 4-B grade in the same school and from the 4-A and 4-B grades in another school:

TABLE 51

SCHOOL	4-B	4-A GRADE
1 . . . . .	9.6	10.6
2 . . . . .	10.2	11.8
Standards . . . . .	9.6	10.4

It is seen, therefore, that the class reported in Table 49, although above the standard, is below the 4-A grade of the other school, which attained a score of 11.8. A further analysis shows that these four classes have made scores equal to or above the standards, and also that school number 2 scored considerably ahead of school number 1.

After the score for the entire class has been secured and interpreted the scores of individual pupils should be analyzed to ascertain whether or not all pupils are properly placed and are receiving proper instruction. For this purpose the class record sheet showing the score of each pupil should always be kept available for frequent reference by the teacher.

By referring to the record of the class in Table 49, it will be seen that the lowest scores made by any boy or girl are 2 and 8 respectively, and the highest 15 and 16 respectively. The lowest score made by any pupil is a score of 2 made by a boy, S. L., whose chronological age is 10 years, 2 months, and whose mental age as later determined by the Stanford Revision of the Binet-Simon Test is 8 years, 2 months, or an Intelli-

gence Quotient of 80.3. The next lowest score is 6, which is also by a boy, W. B., whose chronological age is 11 years, 3 months, and whose mental age on the Binet-Simon Test is 9 years, 3 months, or an Intelligence Quotient of 82.2. The highest score made by any boy is 15, which was made by A. G., whose chronological age is 10 years, 0 months, and whose mental age is 10 years, 0 months, or an Intelligent Quotient of 100.0; and the highest score of any pupil is 16, which was made by a girl, L. J., whose chronological age is 10 years, 8 months, and whose mental age is 12 years, 1 month, or an Intelligence Quotient of 113.2. It is evident, therefore, that some of these children should receive special attention. The pupils who made the low scores are not benefiting from the class instruction to the extent they should. They should be placed in a special class where more individual or perhaps a different kind of instruction can be given.

The question should likewise be raised in connection with the examination of any class, as to whether the pupils who make the highest scores should be advanced to another grade or be assigned to a faster group. In the group above it would seem that the girl, L. J., at least should be given such consideration.

With the aid of the Trabue Language Scales the teacher can quickly determine the general intelligence of the class as a whole and also the general intelligence of each pupil in her class. If there is a question of doubt about certain pupils, the results from the Trabue Language test can be checked by a more refined measurement, such as the Stanford Revision of the Binet-Simon Test. With such knowledge about the mental ability of her pupils the teacher can classify her pupils so that her instruction can be more effective.

#### OTIS GROUP INTELLIGENCE SCALE

This scale has been devised in response to a wide demand for a test which will determine the general mental ability of

children in large groups. Since the ability to read is required to take this test, it is not applicable to persons with less than 3 or 4 years of schooling. Of this test Dr. Lewis M. Terman says: "With subjects of this much schooling, the Otis Scale probably comes as near testing raw 'brain' power as any system of tests yet devised."

**The Aim.** — The aim of this scale is to determine a pupil's general mental ability. It is expected that the Otis Scale will be used for school purposes, to classify, quickly and efficiently, large groups of children on a basis of their mental ages, in order to meet more adequately their individual needs, and that the Binet-Simon Test and others will be used to supplement this scale in cases which are in doubt, or which call for more refined measurements.

**Description of the Test.** — The Otis Group Intelligence Scale is divided into two forms, A and B, which are different in substance but similar in structure. Each form is in a separate booklet. By this means the same group of children can be examined at different times without a knowledge of the tests affecting the results. The total point score for each is the same. Each form has ten tests, as follows:

NUMBER		TIME LIMIT
Test 1	Following directions	5 minutes
Test 2	Opposites	1½ minutes
Test 3	Disarranged sentences	1½ minutes
Test 4	Proverbs	6 minutes
Test 5	Arithmetic	6 minutes
Test 6	Geometric figures	6 minutes
Test 7	Analogies	3 minutes
Test 8	Similarities test	4 minutes
Test 9	Narrative completion	6 minutes
Test 10	Memory	3 minutes

The scale can be used with children in grades 4 through the high school, and even with university students if desired.

**Giving the Test.** — Any person who is able to teach can,

after a little study, apply these tests with a sufficient degree of accuracy to insure satisfactory results. Before attempting to give the tests, however, the teacher should practice on the instructions given in the Manual of Directions, which should always be available. Each child must have a copy of the scale in booklet form. The instructions for each test are written at the top of the test, but divided from the test by a heavy black line. Too much care cannot be exercised in seeing that the children follow specifically the instructions as outlined.

**Scoring Results.** — An examiner's key on transparent paper is provided, which makes the scoring of the papers a very simple matter. For the scoring of all tests except test 3, the check mark ( $\checkmark$ ) opposite each correct answer can be used. The sum of the number of checks will be the score of the individual on the test.

In scoring test 3, a check should be placed after each correct answer and a cross after each incorrect answer only. No attention need be paid to omitted answers. The score will be the number of correct answers minus the number of incorrect answers; that is, "the number of checks minus the number of crosses." (For more detailed instructions for scoring, see pages 29 and 30 of the Manual of Directions, 1919 edition.) The sum of the scores on each individual test will give the individual's score. This score can be placed on the front page of each individual's test sheet, or it can be transferred to a record sheet on which the name of each child can be written and the score on each test, together with his total score, placed opposite his name.

The score of each pupil can be expressed in terms of first, mental age; second, intelligence quotient; third, percentile rank; fourth, coefficient of brightness.

To date no age norms are available from which a child's mental age can be determined. The author, however, is collecting results from these tests wherever they are given,



and undoubtedly will have such age norms in publication in the very near future.

To secure an age norm for the group tested the examination booklets are arranged according to the exact ages of the children. "To do this it will be necessary to take account of the date of the birthday. The Total Score Norm for the age of 12 years may then be taken to be the average of the Total Scores of all pupils whose ages were between 11 years, no months, and 13 years, no months. The Norm for the age of 12 years, 1 month, may be taken as the average of the Total Scores of all pupils whose ages were between 11 years, 1 month, and 13 years, 1 month, etc. The Mental Age of a pupil may then be seen at a glance by noting the age for which his Total Score is the Norm."

The intelligence quotient of a pupil up to 16 years of age can be secured by dividing his mental age by his chronological age. Beyond 16 years of age, the mental age is divided by 16, for the reason that an individual is practically mature at 16 years of age. For a further discussion on determining the scores, especially the percentile rank, and the coefficient of brightness, reference should be made to pages 32 to 36 of the Manual of Directions.

**Interpreting and Using Results.** — In order to indicate to the teacher how she can determine the general intelligence of her pupils with the use of this test, the following results are given which were secured from two 4-B teachers and two 4-A teachers in a city school system. In all 104 children were tested. After the scores were obtained, the papers were classified according to the age groups, *i.e.* all the papers of children 8 years, 0 months, to 10 years, 0 months, were placed in one group; all the papers of children 9 years, 0 months to 11 years, 0 months, were placed in another group, etc. The average of the scores on the papers in the first group gave the age norm of the 9-year-old children; the average of the scores on the papers in the second group gave the age norm

of the 10 year old children, etc. The normal chronological ages for the fourth grade are 9 and 10 years. The age norms for these two groups in these four classes are 54.4 and 54.6 respectively.

The small number of papers makes these figures only tentative norms. These norms would undoubtedly be changed by a larger number of papers, which are necessary to establish reliable standards.

After the age norms for the different ages are secured the mental ages of the different individual pupils can be obtained by noting the age for which the pupil's total score is the norm. For example, R. T., a 4-A pupil, 10 years, 1 month, made a score of 54. His mental age is, therefore, almost 10 years. Another 4-A pupil, S. H., 10 years, 4 months, made a score of 51. He is, therefore, less than 10 years old mentally.

Until there is an age norm for children at every age including years and months, the exact mental ages cannot be determined. When that information is available to the teacher, she can for all practical purposes determine the mental ages of her children whereby a far better grouping or classification can be secured than on the basis of the chronological ages.

#### HAGGERTY'S INTELLIGENCE EXAMINATIONS:

##### *Delta 1 and Delta 2*

**Aim.** — The purpose of this test is to measure the native ability of groups of pupils in the elementary school in order to group them properly or in a limited way to measure their progress.

**Description of Tests.** — The tests appear in two pamphlets, the one, *Intelligence Examination: Delta 1*, for grades one to three inclusive; and the other, *Intelligence Examination: Delta 2*, for grades three to nine inclusive. *Delta 1* contains the following exercises:

- Exercise 1. Oral Directions
- Exercise 2. Oral Directions

- Exercise 3. Copying Designs
- Exercise 4. Copying Designs
- Exercise 5. Picture Completion
- Exercise 6. Picture Completion
- Exercise 7. Picture Comparison
- Exercise 8. Picture Comparison
- Exercise 9. Symbol Digit
- Exercise 10. Symbol Digit
- Exercise 11. Word Comparison
- Exercise 12. Word Comparison

Exercises 2, 4, 6, 8, 10, and 12 determine the pupil's score; the others are preliminary exercises and are not counted in scoring. Simple instructions are given to the teachers for the different performances under each exercise. The difficulty which small children would encounter in reading or in following complicated instructions is avoided. Delta 2 is an adaptation of the army intelligence tests. It has been used more widely than Delta 1. In addition to the examination of 15,000 school children in the state of Virginia, it has been used extensively in many of the larger city school systems throughout the country. It consists of the following exercises:

- Exercise 1. Sentence Reading
- Exercise 2. Arithmetical Problems
- Exercise 3. Picture Completion
- Exercise 4. Synonym-Antonym
- Exercise 5. Practical Judgment
- Exercise 6. Information

The first 5 performances of Exercises 1 and 2, Delta 2, are given to show the nature of the tests:

#### EXERCISE 1

Directions:

1. Read this question:      Do cats see?.....No Yes  
The right answer is Yes; so a line is drawn under Yes.
2. Read the next question: Is coal white?.....No Yes  
The right answer is No; so a line is drawn under No.

Below are a great many more questions. Read them carefully, one at a time, and draw a line under the right answer. When you are not sure, guess.

- 
- |                             |     |    |
|-----------------------------|-----|----|
| 1. Do dogs run?.....        | Yes | No |
| 2. Can a doll sing?.....    | Yes | No |
| 3. Does the sun shine?..... | Yes | No |
| 4. Do men drink water?..... | Yes | No |
| 5. Are all apples red?..... | Yes | No |

EXERCISE 2

Get the answers to these problems as quickly as you can. Use the side of this page to figure on if you need to.

- 
- SAMPLES. — 1. How many are 5 men and 10 men? Answer (15)
2. If one pencil costs 5 cents, what will  
4 pencils cost?..... Answer (20)
1. How many are 30 men and 7 men?..... Answer ( )
2. A boy had 10 cents and spent 4 cents. How  
many cents had he left?..... Answer ( )
3. If you save \$7 a month for 4 months, how much  
will you save?..... Answer ( )
4. If 24 men are divided into groups of 8, how many  
groups will there be?..... Answer ( )
5. A boy had 12 marbles. He bought 3 more, and  
then lost 6. How many marbles did he have  
left?..... Answer ( )

**Giving the Tests.** — A carefully devised manual of directions is provided by the author which must be in the hands of the teacher and thoroughly understood by her before any attempt is made to give the tests. The instructions in the manual are simple so that no teacher should have any difficulty in applying the tests to her class. Each child in the class must be provided with a test. The entire class can be examined at once.

**Scoring the Results.** — The Manual of Directions also provides explicit instructions for scoring the tests. A Scor-

ing Key is provided for both tests. The score of each pupil is the sum of the scores made on the several items of the test. The maximum score for each test is as follows:

<i>Delta 1</i>		<i>Delta 2</i>	
EXERCISE	MAXIMUM SCORE	EXERCISE	MAXIMUM SCORE
2 . . . . .	10	1 . . . . .	40
4 . . . . .	10	2 . . . . .	20
6 . . . . .	16	3 . . . . .	20
8 . . . . .	20	4 . . . . .	40
10 . . . . .	48	5 . . . . .	16
12 . . . . .	25	6 . . . . .	40
Total . . .	129	Total . . .	176

After the score for each pupil is determined these scores are transferred to a class record sheet and the median class score is determined. The results are given in terms of median scores and age norms for each grade.

**Using the Results.**—These Intelligence Examinations have been used with a sufficiently large group of pupils to insure standards that are exceedingly valuable for comparative purposes. Test Delta 2 has been used with about 20,000 children and Test Delta 1 with 4000 children. The following tentative standards are available:

TABLE 52. — GRADE STANDARDS FOR GENERAL INTELLIGENCE TEST:  
DELTA 1

Grade at end of year . . . . .	1	2	3
Score . . . . .	35	55	70

TABLE 53. — AGE NORMS FOR GENERAL INTELLIGENCE TEST: DELTA 1.

Age . . . . .	7	8	9	10
Score . . . . .	35	50	65	75

TABLE 54. — STANDARD SCORES IN GENERAL INTELLIGENCE EXAMINATION: DELTA 2 FOR EACH OF GRADES 3 TO 9 INCLUSIVE

Grade . . . . .	3	4	5	6	7	8	9
Score . . . . .	40	60	78	96	110	120	130

TABLE 55. — AGE NORMS FOR GENERAL INTELLIGENCE TEST: DELTA 2

Age . . . . .	8	9	10	11	12	13	14	15
Score . . . . .	25	43	55	66	77	87	100	115

TABLE 56. — STANDARD SCORES IN EXERCISES 1 AND 2: DELTA 2, FOR GRADES 3 TO 9 INCLUSIVE

Grade . . . . .	3	4	5	6	7	8	9
Exercise 1 . . . . .	14	20	23	27	30	32	35
Exercise 2 . . . . .	5.0	7.0	9.0	10.5	11.5	13.0	15.0

From the above standards the teacher can tell whether or not her class as a whole measures up to the standard for the grade. She can also determine the mental age of each individual pupil in her class, whether or not they are above or below the standard for the grade. With the aid of such a test the teacher should have no difficulty in determining the native ability of each individual pupil in her class.

*The Group Tests for Grammar Grades*, by Dr. G. M. Whipple,<sup>1</sup> are group intelligence tests similar in many respects to the Otis and Haggerty tests. The purpose of these tests is to select the brighter pupils from grades 4, 5, and 6, but the tests quite fully distribute pupils in these grades according to intelligence. It takes 92 minutes to give the tests to a sixth grade, the entire room being tested at once. But when the papers are scored, the teacher has before her a fair measure of the intelligence of each pupil in the group.

<sup>1</sup> See Bibliography at close of chapter.



## THE STANDARD REVISION OF THE BINET-SIMON TEST

This test is possibly the most accurate instrument so far available with which to determine the native ability of American children. On account of the fact that from 30 to 60 minutes are consumed in examining each pupil and that special training is required by the one applying the test, it cannot be used for the examination of large groups of children in a limited time. After the group tests suggested above have been given, there will always be questions which can be settled only by a more refined measurement such as the Stanford Revision of the Binet-Simon Test. It is not too much to expect that the group intelligence tests will open the way and extend the use of the latter test.

An illustration of how this test will supplement the group tests is given in Table 57. The teachers in two 4-A classes in a city school system examined their children with the Trabue Language Scale. In all 44 children were tested. All of them were also examined by a psychologist with the Stanford Revision of the Binet-Simon Test.

The standard for the 4-A grade on the Trabue Scale is 10.4. Nineteen of the 44 children made a score below this standard. On the Binet-Simon Test all but 10 of these 44 children made an Intelligence Quotient above 80.

Table 57 gives the scores on the Trabue Language Scale and the Binet-Simon Test for those 18 children who scored below the standard for the Trabue Language Scale.

From these figures it is seen that 9 of the 14 children scoring from 8.0 to 10.3 inclusive on the Trabue Scale tested between 80 to 90 Intelligence Quotient or below. In the group between 80 and 90 Intelligence Quotient are found, according to Terman, "those children who would not according to any accepted social standards be considered feeble-minded, but who are nevertheless far enough below the actual average of intelligence among races of western

European descent that they cannot make ordinary school progress or master other intellectual difficulties to which average children are equal." Of the children scoring below 8 on the Trabue Language Scale all except one (Intelligence Quotient 81.8) tested below 80 Intelligence Quotient.

TABLE 57

TRABUE LANGUAGE SCALE	INTELLIGENCE QUOTIENT ON BINET-SIMON TEST
6.3	75.8
6.6	81.8
7.6	75.5
7.6	75.
8.	89.9
8.3	89.9
8.3	81.7
8.6	101.
8.6	66.1
9.	93.1
9.3	77.3
9.5	71.5
9.6	73.3
9.6	90.9
10.3	84.8
10.3	96.9
10.3	98.3
10.3	80.9

In this particular city it is the practice to consider pupils eligible to a special class when they test below 80 Intelligence Quotient. It would seem, therefore, that by having the psychologist test with the Binet-Simon Test the children who scored below 10.3 on the Trabue Language Scale, those children could easily be detected who should receive special instruction.

In the same way the children who test very high on the

Trabue Language Scale should be given an individual examination for further classification.

By combining the use of the Binet-Simon test with such group intelligence tests as the Trabue Language Scale, the Otis Intelligence Scale, the Haggerty Intelligence Examinations: Delta 1 and Delta 2, and the Whipple Group Tests for the Grammar Grades, classroom practice can readily be placed on a more scientific basis. The availability of such tests makes it possible for every teacher to know the mental age of every child.

The training of a teacher in a normal school should include a course in the testing of general intelligence so that every teacher may apply any general intelligence test with a reasonable degree of accuracy. Moreover, every teacher could well afford to spend six weeks in the summer in an institution where such training could be secured.

On account of the detailed instructions necessary for the application of the Stanford Revision test, no person should attempt to give it without having access to the "Measurement of Intelligence" by Terman, or some book like it in which such instructions are given.

## BIBLIOGRAPHY

### REFERENCES

- Terman, Louis M., "The Measurement of Intelligence," Houghton Mifflin Company.
- Trabue, M. R., "Composition-Test Language Scales," Contributions to Education, No. 77, Bureau of Publications, Teachers College, Columbia University, New York City.
- Whipple, G. M., "Group Intelligence Tests," Public School Publishing Co., Bloomington, Illinois.

### TESTS

- Haggerty, M. E., "Standard Educational Tests." Intelligence Examination: Delta 1 and 2, World Book Co., Yonkers-on-Hudson, New York, and Chicago. Prices: Delta 1, \$1.50 per package of

25, scoring key, \$.15; Delta 2, \$1.50 per package of 25, scoring key, \$.10. Manual of Directions, \$.35.

Otis, A. S., "Group Intelligence Scale." World Book Company, Yonkers-on-Hudson, New York, and Chicago. Prices: Form A or B including one record sheet, in packages of 25, \$1.50; examiner's key, \$.25; Manual of Directions, \$.25.

Trabue, M. R., "Trabue Language Scales; Scales B, C, D, E, J, K, L, and M," Bureau of Publications, Teachers College, Columbia University, New York City. Price, each scale \$3 per thousand; \$.40 per hundred; carriage extra.

Whipple, G. M., "Group Tests for Grammar Grades." Public School Publishing Company, Bloomington, Illinois. Price, \$.15 each.

Stanford Revision of the Binet-Simon Test, C. H. Stoelting Company, 3047 Carroll Avenue, Chicago, Illinois. Price \$.08 per copy.

## CHAPTER XI

### STATISTICAL TERMS AND METHODS

THE purpose of this chapter is to give only so much information from the science of statistics as the teacher needs to know in order to administer a test, tabulate the scores, and interpret the results. This will necessitate, also, the explanation of statistical terms sufficiently to enable one to understand such terms when used in the discussion of the measurement of any school subject.

**Securing Comparable Results.** — One decided advantage of a standard test is the possibility of comparing the results with similar results in other rooms, other school systems, or with tentative or fixed standards. Manifestly, such comparisons can be made to advantage only when the tests have been given under similar conditions. The following suggestions may be considered as rules of the game for securing comparable results:

1. In giving a test it is essential that the conditions of the test be kept constant.
2. The directions which accompany a test should be followed in every detail. If possible, use a stop watch to secure exact time, when there is a time limit.
3. It is an advantage if the examiner has a clear conception of the nature of the test, its purpose, and the use to be made of it.
4. At the time of giving the test all needed secondary data should be secured, such as name, age, grade, school, date, etc.
5. Most tests as a part of their instructions provide for a preliminary trial in order to make pupils familiar with the

test. In case such provision is not made in the instructions, the teacher should devise a preliminary test which should be similar but somewhat easier than the one to be given, in order that the pupils may thoroughly understand what is to be done and how to do it.

6. The test should be handled as nearly as possible just as any other regular lesson. An appeal to extra effort is allowable, but other comments likely to secure results that are not normal should be avoided. Appeals which are made to the child's desire to do well in the test should be included as a part of the regular instructions, in order that conditions of the test may be uniform for comparisons.

7. For many purposes a single test is sufficient. In case the decision to be based upon a test is of unusual importance, at least two specimens should be taken, or two tests given, or the judging be done by at least two competent judges. In case there is a decided discrepancy between the two results, the teacher will realize that further testing should be done. As will be apparent from further study of statistical methods, a score for a class is much more reliable than for an individual, and the score for an entire city more reliable than that for a single class. This is due to the fact that slight errors tend to balance each other in such a way as to give a more accurate judgment on a large group than on a small group or a single individual.

8. Care should be taken not to use the material of standardized tests for practice purposes.

9. In case the test is given frequently the results will be much more representative if an alternative test of equal value has been provided by the person who devised the test.

**Using a Standardized Test.** — Teachers to-day can scarcely attend an educational meeting or read an educational magazine without hearing about scales and standardized tests, and their advantages in measuring the work of the schools. For the teacher thus interested, but who has not had a normal school



or college course in educational measurement, the following directions are given with the assurance that an intelligent teacher may go forward in such work even though she may not have the help and guidance of a trained supervisor.

1. *Selecting the Test.* — In selecting the test to use, the teacher may well be guided by the particular purpose which she has in mind. The preceding chapters dealing with the available tests in different subjects will permit the teacher to make a choice on the basis of the best test for the particular purpose. In general, those tests should be chosen which have been most widely used, and which require the least time for giving and marking the papers. Yet this is not an infallible rule. The Woody tests are certainly much more valuable than the Courtis tests in arithmetic. Yet the Woody tests have been given very little in comparison with the wide use of the Courtis tests. They are a little more difficult to administer and to score. Yet, from the standpoint of value in diagnosing the pupils' difficulties, they are far superior to the Courtis tests. The tests that are going to survive and show value in the next few years cannot be determined at this time. The final judgment upon the test must be passed by the teacher in the schoolroom on the basis of its value in helping her in her work of discovering the needs of the children and applying the appropriate remedies. It may be properly assumed that, although a test is more difficult and requires a longer time, if it is superior in every respect, the teacher will find the time for giving it. It requires considerable time to give Gray's Oral Reading Test, yet the results of giving the test are so valuable that the teacher does not hesitate to take the necessary time for giving it.

When a test may be chosen on the basis of difficulty, as in the use of the Ayres Spelling Scale, the teacher should keep in mind that a good test should be so difficult that no pupil will make a perfect score, and sufficiently easy that most pupils in the grade will secure a score which is reasonably satisfactory.

2. *Giving the Test.* — In giving a test the teacher should follow carefully the printed directions which accompany the test. This is the chief rule to keep in mind. Other details are mentioned above under "Securing Comparable Results." The teacher who has time and is willing to experiment may easily demonstrate the possibility of changing a score by a slight change in directions or by a different attitude in presenting the work to children. The chief consideration, if comparisons are to be made, is that the attitude, detailed directions, and every element entering into the giving of the test shall be as indicated in the directions, so that pupils in one city, or even in one state, may be compared with those in another. In handwriting, for example, pupils should be so instructed and handled that they will write at their natural rate, thus securing results in the test that will represent the normal situation.

3. *Scoring the Papers.* — Every test provides printed directions for scoring the papers in order to aid teachers in securing uniformity of results. These printed directions should be followed implicitly. If the teacher has opinions as to what should be done, and these opinions are different from the directions, such opinions should be abandoned if the results of the test are to be used for comparative purposes.

The teacher is urged to have the pupils aid in scoring the papers in so far as it is possible. This can be done very largely in arithmetic, in spelling, in certain reading tests, and, to an extent, in writing. The chief purpose of involving the child in the grading is to further increase his interest. This is an incentive and a motive which is worth while for teaching purposes, and which will lead the child to greater effort in order that he may score higher in a future test.

4. *Tabulating Results.* — Directions for tabulating results or distributing the scores are provided in connection with most of the tests. A common method of making a distribution is to arrange the papers in order. The teacher can

then draw off the scores, noting the number of papers falling at each point. This gives the distribution. For further use, the teacher will need to supply the names of pupils opposite each score, or, in case she is noting mistakes, opposite each mistake. The results of any test cannot be intelligently used until they have been arranged in some systematic order, particularly if the number of pupils involved is large.

5. *Statistical Calculations and Graphic Representations.* — The statistical points to be determined after the results of a test have been tabulated are usually the median, the quartiles, and, sometimes, the average or standard deviation. These points will be explained in the next section of this chapter. When understood, they are very valuable, particularly in making a comparison of one room or one school system with another.

To represent the scores graphically often helps the teacher to see points which would otherwise remain hidden. A graphic representation is made by noting the number of scores falling at each point of the scale, and representing the number by the distance from the base line, and then drawing a line connecting all of these points. The height of the line above the base line enables the teacher to see at a glance just what is happening in her class.

6. *Interpretation of Results.* — The teacher is warned to avoid conclusions until she has mastered the technique and the significance of the test and has given it to different groups, or enough times to the same group, to clear up in her own mind the various questions that may arise in giving the test. Especial care should be taken not to draw far-reaching, general conclusions from a test. A test is usually devised for a specific purpose. The significance of the test in other fields can be known only through the figuring of coefficients of correlation after a large number of cases has accumulated. A good drawing of the moon by an eighth grade pupil means a good drawing of the moon — not an artist, not an astronomer.

Nothing should be taken for granted. Mistakes will be avoided by caution, and fear will be eliminated by a thorough understanding.

7. *Applying Remedies.* — The ultimate purpose of a test, so far as the individual teacher is concerned, is to enable her to see the needs of her pupils and to search out the appropriate remedies. The discovery of the remedy in any subject takes her into the question of methods of teaching, but this is a desirable result. To use a test for measurement only, without carrying the work forward to a point of use and application in better teaching, is to close the eyes to the significance of a situation after it has been revealed. The teacher, after giving a test, is in the position of a specialist who has diagnosed a bodily ailment. The diagnosis means nothing unless the appropriate remedy is applied. The recognition of this fact leads a teacher or a group of teachers, again and again, into the study of methods of teaching with reference to the subject tested.

8. *Coöperation.* — In a city system, the closest possible coöperation is urged between supervisor and teachers, not only for the benefit of the teachers, but as well for the benefit of the supervisor. Coöperation, understanding, and mutual confidence are always valuable assets, and especially so in the use of the tests which may reveal teacher weaknesses as well as pupil weaknesses. The teacher, however, will be the first to want to correct any revealed defects, and her interest and coöperation will enable the superintendent or supervisor to secure other important results, such as :

- a. A more scientific attitude toward school work.
- b. A closer checking of results and a realization that pupil errors are specific and need individual attention.
- c. Better time allotments, more definite assignments, a clearer conception of the objectives to be attained, and more efficient methods of teaching.

**Statistical Terms.** — The purpose of the statistical treat-

ment of scores is intelligent interpretation. The first step in the handling of scores is to give them systematic arrangement.

A *distribution* is a systematic arrangement of scores.

A *table of frequency* is a table showing the scale and the distribution of scores at each point on the scale.

The following are the unarranged grades of seventy-seven sixth grade pupils in arithmetic: 74, 92, 65, 69, 76, 80, 62, 73, 85, 81, 79, 66, 59, 75, 76, 81, 84, 74, 55, 73, 86, 75, 71, 60, 92, 85, 76, 82, 50, 65, 92, 100, 81, 75, 85, 97, 65, 91, 85, 86, 72, 55, 75, 75, 72, 77, 62, 95, 87, 75, 75, 70, 76, 87, 85, 82, 67, 90, 81, 95, 80, 86, 80, 75, 67, 70, 72, 84, 76, 70, 88, 72, 80, 75, 67, 82, 72.

Thus arranged, the scores have little significance. They need statistical interpretation. The following *table of frequency* shows a scale with intervals of 5, and on the right hand side the number of scores at each point on the scale. This right-hand column represents the *distribution*. A grade is recorded at the nearest "5" point on the scale, thus, 74 is recorded or scored at 75, 92 at 90, etc.

TABLE 58. — FREQUENCY TABLE: SHOWING THE ARITHMETIC SCORES OF 77 SIXTH GRADE PUPILS

GRADES, INTERVALS OF 5 (SCALE)	NUMBER OF SCORES AT EACH POINT ON SCALE (FREQUENCY)
50	1
55	2
60	4
65	7
70	10
75	19
80	12
85	12
90	6
95	3
100	1
Total	77



This table is much more useful than the undistributed grades, as it enables the teacher to see the number of pupils (or number of scores) at each point on the scale.

Special significance is usually attached to certain points on the scale, such, for instance, as the *passing mark*. If 70 is the passing grade, the teacher sees at once that 14 of the pupils have failed.

Other points on the scale that have statistical value are the median, the quartiles, the mode, the average, and the range.

The *median* is the middle score, or the point on the scale above and below which an equal number of scores fall after the scores have been arranged into a table of frequency. In Table 58 there are 77 scores, so that the middle one would be the 39th score from either end of the distribution. The 39th score falls at 75, and therefore 75 is the median score. In case of an even number of scores, the median is the average of the two middle scores.

The *quartiles* are the points on the scale arrived at by taking  $\frac{1}{4}$  and  $\frac{3}{4}$  the scores, counting in from either end. It is usual to start at the bottom of the scale, so that counting up until  $\frac{1}{4}$  the scores have been covered locates the point on the scale known as the *first quartile*, and the distance up the scale necessary to include  $\frac{3}{4}$  of the scores locates the *third quartile*. The second quartile is seldom referred to as it is the same as the median. It is evident that the first and third quartiles are the points midway between the median and the extremes. The *middle 50%* is a term frequently used. It represents the number of scores falling between the first and third quartiles.

The *extremes* are the outside limits of the distribution, and the distance between the extremes indicates the *range* of the distribution.

The *mode* is the point on the scale where the greatest number of scores fall. In Table 58, the mode is 75.

The *average* is found by adding the scores together, and dividing the sum by the number of scores.



*Deviation.* — Some method of indicating by a single figure the deviation of the scores from some central point like the median is frequently used. *Average deviation* is most often used and it is found by taking the average of the deviations of the individual scores from some central tendency, usually the median. *Standard deviation* is also used to express deviation. It equals the square root of the sum of the squares of the deviations from the arithmetical average (although the median may be used instead of the average).

The teacher will have little use for figuring deviation in the present volume.

*Correlation.* — The relation between two paired series may be expressed by a single figure known as the coefficient of correlation. The figure ranges from  $-1$  to  $+1$ , the latter figure representing perfect correlation or agreement. The work of the present volume will not require the derivation of this term, so the reader is referred to works listed in the bibliography in case of a desire to know the method of figuring the coefficient of correlation.

#### BRIEF BIBLIOGRAPHY

- King, W. I., "Elements of Statistical Method," The Macmillan Company, New York, 1915.  
Thorndike, E. L., "Mental and Social Measurements," Teachers College, New York, 1913.  
Brinton, W. C., "Graphic Methods for Presenting Facts," Engineering Magazine Co., New York, 1914.  
Rugg, H. O., "Statistical Methods Applied to Education," Houghton Mifflin Company, Boston, 1917.  
Buckingham, B. R., "Statistical Terms and Methods," Seventeenth Yearbook of the National Society for the Study of Education, Chap. IX, Part II, 1918.

## CHAPTER XII

### THE TEACHERS' USE OF SCALES AND STANDARDIZED TESTS

THE college instructor blames the high school teacher, the high school teacher complains of the grade teacher, each grade teacher above the first grade finds fault with the poor work of the teacher in the grade below, and the first grade teacher in turn is chagrined at the shortcomings of the home training. Must this go on indefinitely? Whose opinion should prevail? Is it not possible to get away from personal opinion to an agreed-upon consensus of opinion? May we not replace the constantly conflicting subjective standards with definitely defined objective standards?

**Present Grading System.** — If 20 mechanics were sent out into a mill yard to cut and bring back a steel rod just long enough to reach from one girder to another, but were not given the measured distance between the girders before going, nor permitted to take a ruler or tape to use in selecting the rods, no experiment is needed to prove that each one of the 20 rods would be different in length and no one of them would exactly span the distance from girder to girder except by chance. On the other hand if the foreman were to use a steel tape in measuring the width between the girders, and were to permit the mechanics to measure the length of the rods before cutting them, they would return with 20 rods each meeting with his approval.

Is it possible for the school foreman, the teacher, to replace her subjective standard, her mere opinion, by an objective standard approximating the steel tape of the shop? The need of more accurate, objective standards in grading is

generally appreciated. The following are some of the evidences of such need :

(1) *There are constant complaints from teachers in upper grades, as indicated above, against the poor quality of work done in the lower grades.*

(2) There is wide variation in the distribution of grades among the various departments of the same school. In one high school, for example, 80% of the English grades were 90 or above, while only 4% of the mathematics grades were 90 or above. In the same high school, the German teacher gave 70% of her pupils 90 or above, while the Latin teacher gave only 2½% of her pupils a grade of 90 or above.

A recent study of college grading well illustrates this point. The study covered a total of 12,782 grades by 10 professors covering a period of 5 years. The grades given by professors number 1, 3, and 4 are shown herewith :

PROFESSOR	GRADES (Total)	FAILED	75-79	80-84	85-89	90-94	95-100
No. 1 . . .	1071	32.1%	12.7%	15.9%	14.9%	12.7%	11.5%
No. 3 . . .	1422	9.8	7.0	11.9	15.9	14.4	40.7
No. 4 . . .	2196	3.3	6.2	19.3	36.2	28.5	6.3

The contrast between Professor No. 1 and No. 3, who represent the extremes, is brought out more strongly by the graphic representation (see Fig. 16) than by the table. Professor No. 1 fails approximately one third of his students and then distributes the others about equally among the 5 remaining points of the scale. Quite the opposite, Professor No. 3 gives two fifths of his students an honor grade, and then distributes the other grades about equally among the 5 other points of the scale. These figures are in the main true for each of the 5 years studied, without regard to the maturity of the students, whether they be freshmen, sophomores, juniors, or seniors.

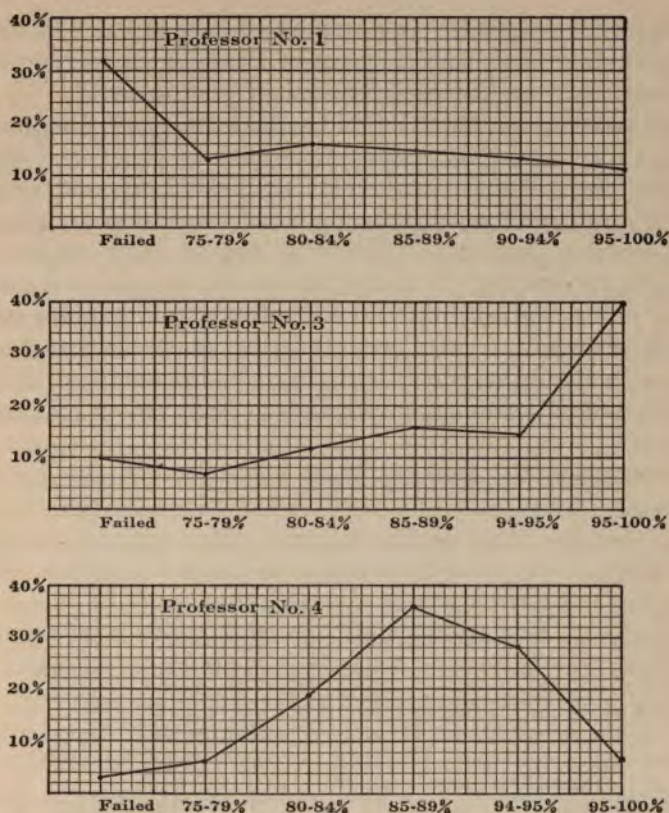


FIG. 16. — Showing graphically the distribution of grades given by three college professors at Iowa State College.

A study of the distribution of the grades given by the faculty of any large high school or college is likely to show similar results, unless the problem of grading has received special attention.

(3) There is a wide variation in the distribution of grades among teachers of the same department. Of 2 instructors in the same department 1 gave to 43% of his students the

grade of "excellent" and to none the grade of "failure," whereas the other gave to none of his students the grade of "excellent" and to 14% the grade of "failure."<sup>1</sup> There must have been a few good and a few bad in each group.

(4) The fact that pupils transferring from one school system to another are frequently demoted indicates that minor details rather than large fundamental considerations are the determining factors in classifying them. Since pupils are constantly shifting, in many schools as high as 20% being new to the system each year,<sup>2</sup> this is a very important item. In fairness to the child, as well as the school from which he came, it should be possible to determine his standing through the use of objective standards, and so place him in the proper grade.<sup>3</sup>

**Differences in Grading Same Paper.**—A study by Dr. Daniel Starch illustrates very clearly the variation among teachers of a single subject in grading that subject. A paper in English was submitted to 142 teachers of English. The grades varied from 50 to 97, the passing grade being 75. Twenty-six of these teachers, or 18%, marked the paper a failure, that is, graded it below 75. On the other hand, 14

<sup>1</sup> Starch, Daniel: "Educational Measurement," p. 3, quoted from Dearborn.

<sup>2</sup> Typical facts with reference to the proportion of school children who leave school, because of leaving the city, are easily gathered from current school reports. The following are illustrative:

In Waterbury, Connecticut, 1914-15, there was a total enrollment of 13,954. Of this number 903, or 5.4 per cent, left school during the year. Of those who left school, 426, or 47.2 per cent, left the city. Similar facts for other cities follow:

In Des Moines, Iowa, 1913-14, 10.7 per cent left school. The proportion of those who left the city was 01.3 per cent.

In Decatur, Illinois, 1913-14, 15.7 per cent left school. Of these 62.7 per cent left the city.

In Connersville, Indiana, 1910-12, two years combined, 14.8 per cent left school. Of these 57.8 per cent left the city.

Pupils who leave the city will usually enter other school systems.

<sup>3</sup> Asst. Supt. O'Hara in the May, 1916, number of *Elementary School Journal* calls attention to the value of standard tests for placing new pupils in the right grades.



of the group marked it 90 or above, indicating that in their opinion it was a very superior paper.

In mathematics, a similar test gave results that were even more surprising, particularly so in view of the fact that mathematics is considered one of the exact sciences. A geometry paper which was submitted to 118 teachers received grades ranging from 29 to 92, the passing mark being 75. Sixty-eight of the teachers, or nearly 58% of them, marked the paper a failure. Fifty of the group marked it 75 or above, one giving it a grade of 92.

A history paper graded by 70 teachers showed similar variations, the grades ranging from 43 to 90.

This but illustrates the present chaos resulting from the lack of standards in grading an ordinary examination paper. When this is multiplied by the variation in sets of examination questions, it is apparent that on the present basis of examinations it is absolutely impossible to compare one system with another, one grade with another, or to compare from month to month the same grade with itself. It is unnecessary to discuss fully the above points. Others might be added, all indicating the need of objective standards.

**Uniform Examination Not Satisfactory.** — One may ask if the purposes of an objective standard for measuring school achievement cannot be accomplished by a uniform course of study, uniform examination questions, and uniform grading. These items may properly receive attention in order. In the first place a uniform course of study is undesirable. It must be adjusted to community demands and pupil interests. It should differ greatly for children from the exclusive residence districts of New York City, and the children from Iowa farm homes. To attempt to secure a rigid uniformity in the course of study would be deadening in the extreme. The course of study should be flexible and provide for local variations. To possess knowledge which is useful and usable is much more fundamental in a democracy than to strive for a large common



intellectual possession composed too largely of material which is stale and useless.

In the second place, all will agree that there is nothing more baneful and stupefying in its influence than a rigid examination system. It makes subject matter the aim and end. It leads to cramming. It militates against use and application. It directs pupils to words in books instead of to life's real problems and their solutions. And strange as it may seem, a standard test will accomplish the desirable results of a uniform examination system without the undesirable results appearing. In subjects for which standard tests are not available, examinations must continue to be used. They have a value if rightly used.

In the third place, all will admit that uniformity of grading is desirable. It is difficult, however, with an ordinary examination, although common practice may be improved by adopting a *x*-point system and distributing grades according to the normal curve of distribution. How to improve the grading of a group of teachers along these lines has been well explained by Gray, Meyer, Dearborn, Judd, Starch, Kelly,<sup>1</sup> and others. But one of the greatest advantages of the standard test or scale is that it greatly aids in securing uniformity of results in grading. In order to standardize a test, specific directions have of necessity been prepared for giving the test and for grading the returns. All of this means greater uniformity in grading, and greater fairness to individual classes or pupils in case of comparison. In fact, a standard test in this way, a well-constructed uniform examination, accompanied by specific directions which greatly aid in securing uniformity and fairness.

**Standard Tests.** But a standard test is much more than

<sup>1</sup> Kelly, C. C. "Standard Tests and Their Distribution." Contributions of the Bureau of Educational Research, Columbia University. This volume contains a good bibliography on the study of school and college tests and grading.

a uniform examination. The standardization of a single test or scale often requires a year or more of intensive work by one of our ablest educators. Not only must the subject matter be carefully selected and adapted to pupil ability, but it must be tried out with thousands of pupils, revised, and again tried out, until every detail of the test, its administration, its evaluation, and the grade or age standards, has been determined. Such an undertaking is too much to expect from the overworked teacher. But the teacher may properly be expected to profit by the standard tests of subject matter which have become available.

The difference between an examination and a standard test, as well as the progress of measurement in education, is fairly well illustrated in the attempt to measure arithmetic in the two Cleveland surveys, the first by a local commission in 1906, the second by a survey committee composed of educational experts selected from all parts of the country only nine years later, 1915.

The arithmetic test given in the first Cleveland survey was devised by men of maturity and judgment, but had not been standardized. It was not even based upon a wise selection of subject-matter, and it could not lead to any valid conclusions. It was used in at least one later survey.<sup>1</sup> It did not justify further use, although it was doubtless as good as any test that could have been quickly devised under the circumstances. At the time Thorndike's writing scale had just appeared but had not come into general use, and there were no standard tests.

In 1915, however, the work of the Cleveland schools was *measured* in a scientific manner which carried conviction everywhere. In writing, spelling, arithmetic, and reading, scales or standard tests were applied which clearly revealed the grade to grade progress of the pupils, made possible comparison of one building with another, and permitted comparison of the

<sup>1</sup> East Orange, N. J., 1911, by Dr. E. C. Moore.

work in Cleveland with similar work in other cities throughout the country.

While a particular teacher need not be greatly concerned about having a test that will permit comparison of the work in one city with the work in another, or even a comparison of her work with the work of other teachers in the same grade throughout the system in which she works, yet she should be concerned about the progress of the children within her own room. She should know the results of her work. She should have a device for the definite measurement of progress, due to a particular method, or a given time devoted to the work. These aims cannot be accomplished through the ordinary examination. They can be accomplished only through the use of scales and standardized tests.

**Initiating the Use of Standard Tests.**—Whether the initiative in the use of standard tests be taken by the teacher, the superintendent, or a survey commission, the final result should be to help the teacher, and through her, the pupil.

Miss Laura Zirbes,<sup>1</sup> of the Cleveland University School, took the initiative in the use of standard tests, completely transformed her own theory and practice, and brought new life and more rapid progress to her pupils. In Boston, the initiative came from the central office, but in such sympathetic and coöperative form that teachers were effectively reached. Of more pronounced effect probably than any of these factors, however, was the stimulation among the Boston teachers of an inquiring attitude towards the whole problem of arithmetic instruction. "The results from the tests have shown the need of improvement; they have shown that the problem of arithmetic teaching is not yet solved, and they have prompted many teachers to study their own work as the first step towards improving methods of instruction."<sup>2</sup> Later

<sup>1</sup> "Diagnostic Measurement as a Basis of Procedure," *Elementary School Journal*, March, 1918, pages 505-522.

<sup>2</sup> Boston, Educational Bulletin No. X.

an entire bulletin<sup>1</sup> was devoted to showing teachers and principals how to use the results of standard tests in reaching individual pupils and improving instruction.

The teacher who uses a standard test in her own room for the purpose of knowing her pupils or locating the weak places in her instruction may take pride in the fact that she is putting herself in line with a vast army of scientific workers in education. She determines the median ability of 30 pupils in a single grade, the distribution of ability, the points of weakness, and the remedies to apply. A principal does the same for the entire building; the superintendent for the entire school system; a state bureau for the state; and a research specialist, by combining city and state results, gets norms of performance for a nation. The teacher thus sees herself as a contributor in a great piece of constructive work in scientific education, and she may, if she wishes, locate her particular group of children with reference to the thousands of other children throughout the county, — she may feel the thrill of being one of the 700,000 lieutenants who marshal the army of 22,000,000 American school children, in the interests of a safer and saner democracy.

**Uses of a Standard Test.** — However, the most helpful point for the present purpose is that standard tests should be used by the individual teacher for the purpose of finding the weaknesses in her own work, evaluating methods, and definitely measuring the progress of her own pupils. It will be worth while to enumerate in order the uses that a teacher may make of a standard test. Some of these are in common with the uses which may be made of the results of standard tests by principals and superintendents, but many of them apply directly to the particular schoolroom and are in addition to other uses. Standardized tests may be used:

1. To determine conclusively whether or not a pupil is making progress.

<sup>1</sup> Boston, Educational Bulletin No. XIII.

2. To determine how much progress a pupil has made in a given time.

3. To determine whether a pupil should be promoted, retained, or reclassified, in so far as the mastery of subject matter is made a condition of progress. Dr. Starch states that promotion on the basis of measured ability would save one year to one third of the pupils in the public schools.<sup>1</sup>

4. To determine even more accurately whether or not the class is making progress and the amount of such progress.

5. To determine whether or not a class is up to standard when received from another teacher. This use of the standard test would remove the constant complaint of teachers that the work has not been covered in the preceding grades.

6. To justify a year's work with a class on the basis of actually measured progress. This will make it possible to show to a prejudiced principal or superintendent (if such there be) that reasonable progress has been made by a class.

7. To show results in a manner that completely discounts the advantages of another teacher more attractive and popular, in case such teacher depends upon winning promotion by methods not contributing to pupil progress.

8. To detect the fact, in case more time cannot profitably be spent with retarded pupils. See, for example, the conclusion of Superintendent Bliss of Montclair, N. J., that a group of subnormal pupils could not profit by further work in arithmetic.<sup>2</sup>

9. To release bright pupils from further work after determined standards have been reached, as long as said standards are maintained. The teacher would thus limit the work required along mechanical and routine lines. Rice's articles<sup>3</sup> on the "Spelling Grind" over 20 years ago emphasized the fact of wasted youth through the schools. Over emphasis

<sup>1</sup> Fifteenth Yearbook of the National Society for the Study of Education, Part I, p. 146.

<sup>2</sup> *Ibid.*, p. 75.

<sup>3</sup> *The Forum*, XXIII: 163-172, 409-419.



upon the mechanical phases of school work closes the door to story, romance, history, literature, music, and play.

10. To test one method against another by the amount of measured progress made by the pupils, *e.g.* textbook procedure versus large motivated problems, as a basis for developing ability in solving reasoning problems (in so far as devised tests adequately measure this educational product). It is apparent that such use of standardized tests would replace the trial and error method as a means of determining correct procedure, and would replace it by a method much more scientific.

11. To test one class plan, study plan,<sup>1</sup> or administrative device against another, by measured results with the pupils.

12. To determine the proper apportionment of school time to various subjects of study and other school activities. This use of standard tests has been well pointed out by Dr. Haggerty.<sup>2</sup>

**Standard Test Saves Time.** — Naturally the teacher asks, "But will not this scientific testing require a much larger time expenditure than I can give to it? I'm crowded for time as it is."

This question can be answered only on the basis of the experience of other teachers. That experience shows that after the technique is once mastered, the time required for standard testing is not more, but frequently less, than the time consumed in marking papers under the old examination system. After the writing scale has been used for a while, has been conveniently posted for reference by pupils, and has been explained to them, the teacher will find that a committee of pupils can be relied upon to grade the writing of the room, honestly and quite accurately. In fact each pupil will grade

<sup>1</sup> See p. 113, *Schoolman's Week Proceedings* (University of Pennsylvania), April, 1918, for comparison of class study and independent study in spelling. Reported by J. N. Adee, Superintendent Schools, Johnstown, Pa.

<sup>2</sup> *School and Society*, IV: 761-771.



his own writing by comparison with the scale. After the spelling test has been given, pupils may be allowed to exchange papers and correct them while the teacher gives the correct spelling of the words. Likewise in arithmetic, the pupils can help the teacher in quickly grading the papers. This help by pupils in the simpler tests should be encouraged not alone because it saves the time of the teacher, but chiefly because it creates a desirable interest and stimulates the pupils to put forth a greater effort to reach a given standard.

**Standard Test a More Effective Tool.**—The question with regard to the time required for giving standard tests is a legitimate one, and an effort has been made to answer it. Every conscientious teacher will agree, however, that time is not the chief consideration. She puts in a full quota of time each day, and will continue to do so. If she is as wise as conscientious, she will also provide time for sufficient sleep and recreation each day. The chief consideration is that the teacher in mastering the details of the use and interpretation of a standard test is equipping herself with a more effective tool of service. Why should the teacher guess and estimate when she can measure? The unsatisfactory nature of the present grading system has been dwelt upon. A grade of 85 in one room cannot be compared with a grade of 85 in another room. The present unscientific method of grading must be replaced by scientific procedure if we are to continue to make educational progress. Improvement is certainly hampered by the use of a system which does not even permit of comparison, and thus give a definite measure of progress. Under the old system when two schools determined to compare the spelling ability of their pupils, all that they could do was to get the pupils together and have them compete in a spelling match. And yet as we look back upon the spelling match we see that the result was finally determined by the one best speller. The general merit of spelling in one school as compared with the general merit in the other was not determined.

**Measuring a Human Product.** — The teacher may insist that she is dealing with a delicate human product. This is true; and yet, as Thorndike has pointed out, mental products can be measured and are being measured. "Whatever exists, exists in some amount." The work of the physician probably compares as closely as any other with that of the teacher. We want a physician who is kind and sympathetic, but we are not willing that these qualities be substituted for accurate and adequate knowledge. Regardless of his kindness and sympathy, he counts the pulse, and takes the temperature. In case an anæsthetic is to be administered, he calls in an expert to determine the amount and to administer it according to standard methods. In case of a surgical operation he again calls for an expert, frequently a busy, unsympathetic stranger. In all of this work, regardless of his kindness, sympathy, geniality, and his spiritual qualities in general, he relies upon accurate knowledge, definite measurement, and tested skill. He proceeds scientifically. The teacher should do likewise.

"It is a popular superstition that human action, personality, and behavior will be penned up and hindered when measured by logical categories and fixed units. But, just as the pound weight has not interfered with the production of butter, and the yardstick has not obstructed improvement in the manufacture of cotton or other goods, so methods of teaching, it may be assumed, 'will improve and develop freely, even when fixed standards are applied.' The spirit can still go where it listeth. Measurement must meekly follow, gather up the results, and give them a value."

Weights and measures call to mind definite units, such as pound, quart, and yard, and these are infinitely more valuable for commercial purposes than "as much as a man can lift," "a small jar full," or "the length of a man's arm." Standards have made commercial transactions possible at great distances on a basis of perfect understanding and fairness.

There is no doubt that teaching and the products of school work are going to be benefited in a similar manner by the application of definite standards of measurement. Measurement is always taking place in one form or another. School work is being constantly noted as *good*, *fair*, or *poor*, as *satisfactory* or *unsatisfactory*, and is constantly being rated by such standards as are available, be these standards crude or otherwise.

Many large cities have established bureaus of measurement and efficiency. Each bureau has a head with an adequate clerical staff. Such an organization is needed in a large city even when the teachers administer and grade the tests. A central bureau can establish city standards, make valuable comparisons, and interpret results in a way to be most valuable and helpful to all teachers as well as superintendents and supervisors. But more and more the directors of central bureaus realize that they are failing unless they reach the individual teachers. Dr. Ballou emphasizes this on every page of his recent bulletin interpreting results in arithmetic.<sup>1</sup> He assures us that in the last analysis "the teacher must find out what her trouble is and then apply the remedy."

The present work makes no effort to discuss the complete list of available tests, but instead is limited to such tests as have been standardized sufficiently to recommend their use to the teacher who, for the most part, is untrained in the use of statistical methods. In beginning the work in measurement, teachers should make no effort to employ all available tests, but should carefully select the test to be given. As pointed out by Ballou, teachers will do well to give tests that are reasonably simple, that can be scored and tabulated with reasonable ease, and that have been given to a sufficient number of children so that well-founded standards of achievement have been established, the first assumption always being that the test measures desirable phases of school products.

<sup>1</sup> Boston, Educational Bulletin No. XIII.

# BIBLIOGRAPHY

1. "Standards and Tests for the Measurement of the Efficiency of Schools and School Systems," Part I, Fifteenth Yearbook of the National Society for the Study of Education (1916), Public School Publishing Co., Bloomington, Illinois.
2. "The Measurement of Educational Products," Part II, Seventeenth Yearbook of the National Society for the Study of Education (1918), Public School Publishing Co., Bloomington, Illinois.
3. Indiana University Studies, by Haggerty, M. E.: 27, "Arithmetic: A Coöperative Study in Educational Measurements"; 32, "Studies in Arithmetic."
4. Ayres, L. P., "A Survey of School Surveys," Indiana University, Second Conference on Educational Measurements, pp. 172-181.
5. Ballou, Frank W., "Improving Instruction Through Educational Measurement," *Proceedings N. E. A.*, 1916: 1086-1093.
6. Ballou, Frank W., *Bulletins of the Department of Educational Investigation and Measurement*, Boston, Nos. X, XIII.
7. Bobbitt, J. F., Twelfth Yearbook of the National Society for the Study of Education, Part I, pp. 7-96.
8. Curtis, S. A., "Standardization of Teachers' Examinations," *Proceedings N. E. A.*, 1916: 1078-1086.
9. Cubberley, E. P., "The Significance of Educational Measurements," Indiana University, Third Conference on Educational Measurement, pp. 6-20.
10. Freeman, Frank W., "Some Practical Studies of Handwriting," *Elementary School Journal*, 14: 167-179, December, 1913.
11. Haggerty, M. E., "Some Uses of Educational Measurements," *School and Society*, 4: 761-771, November 18, 1916.
12. Harlan, Chas. L., "A Comparison of the Writing, Spelling, and Arithmetic Abilities of Country and City Children."
13. Judd, Chas. H., "Reading Tests," *Proceedings N. E. A.*, 1915: 561-565.
14. Judd, Chas. H., "Standardized Units of Achievement of Pupils and Measurable Standards of School Administration," *Proceedings N. E. A.*, 1917: 721-724.
15. Judd, Chas. H., "Measuring the Work of the Public Schools," Survey Committee of the Cleveland Foundation, Cleveland, Ohio.
16. Judd, Chas. H., "A Look Forward," Seventeenth Yearbook of the National Society for the Study of Education, Part I, pp. 152-160.

17. Melcher, George, "The Two Phases of Educational Research and Efficiency in the Public Schools," *Proceedings N. E. A.*, 1916: 1073-1078.
18. Moore, E. C., Report on East Orange, New Jersey, 1911.
19. Morrison, J. Cayce, "The Supervisor's Use of Standard Tests of Efficiency," *Elementary School Journal*, 17: 335, January, 1917.
20. O'Hern, Joseph P., "Practical Application of Standard Tests in Spelling, Language, and Arithmetic," *Elementary School Journal*, 18: 662-679, May, 1918.
21. Rice, J. M., "The Futility of the Spelling Grind," *Forum*, 23: 163-172 and 409-419.
22. Starch, Daniel, "Educational Measurements."
23. Strayer, George D., "The Use of Tests and Scales of Measurement in the Administration of Schools," *Proceedings N. E. A.*, 1915: 579-582.
24. Strayer, George D., *et al.*, "Report of the Committee on Tests and Standards of Efficiency in Schools and School Systems," *Proceedings N. E. A.*, 1913, 392-406.
25. Thorndike, E. L., "The Elimination of Pupils from School," Bulletin No. 4, 1907, United States Bureau of Education.
26. Wilson, G. M., "The Handwriting of School Children," *Elementary School Teacher*, 11: 540-543, June, 1911.
27. Wood, Ernest R., "Tests in Efficiency in Arithmetic," *Elementary School Journal*, 17: 446-453, February, 1917.
28. Zirbes, Laura, "Diagnostic Measurement as a Basis for Procedure," *Elementary School Journal*, 18: 505, March, 1918.

## APPENDIX

DATA FOR ESTIMATING THE DEGREE OF DIFFICULTY REQUIRED TO PRODUCE 20 PER CENT OF WRONG OR OMITTED RESPONSES WHEN A GIVEN STEP OF THE SCALE PRODUCES FROM 8 PER CENT TO 40 PER CENT OF SUCH. THEY ARE USED IN CONNECTION WITH THE THORNDIKE READING TESTS FOR THE DETERMINATION OF PUPILS' SCORES.

GIVEN PERCENTAGE	ADD	GIVEN PERCENTAGE	ADD	GIVEN PERCENTAGE	ADD	GIVEN PERCENTAGE	ADD
8.0	.84	5	.61	13.0	.42	5	.26
1	.83	6	.60	1	.42	6	.25
2	.82	7	.60	2	.41	7	.24
3	.81	8	.59	3	.40	8	.24
4	.80	9	.58	4	.39	9	.23
5	.785	11.0	.57	5	.39	16.0	.23
6	.78	1	.57	6	.38	1	.22
7	.77	2	.56	7	.37	2	.21
8	.76	3	.55	8	.37	3	.21
9	.75	4	.54	9	.36	4	.20
9.0	.74	5	.53	14.0	.36	5	.20
1	.73	6	.52	1	.35	6	.19
2	.72	7	.52	2	.35	7	.18
3	.71	8	.51	3	.34	8	.18
4	.71	9	.51	4	.33	9	.17
5	.70	12.0	.49	5	.33	17.0	.17
6	.69	1	.49	6	.32	1	.16
7	.68	2	.48	7	.31	2	.15
8	.67	3	.48	8	.30	3	.15
9	.66	4	.47	9	.30	4	.14
10.0	.65	5	.46	15.0	.29	5	.14
1	.64	6	.45	1	.28	6	.13
2	.63	7	.45	2	.27	7	.12
3	.62	8	.44	3	.27	8	.12
4	.62	9	.43	4	.26	9	.11



GIVEN PERCENTAGE	ADD	GIVEN PERCENTAGE	SUBTRACT	GIVEN PERCENTAGE	SUBTRACT	GIVEN PERCENTAGE	SUBTRACT
18.0	.11	20.0	.00	8	.19	6	.37
1	.10	1	.00	9	.19	7	.37
2	.10	2	.01	24.0	.20	8	.38
3	.09	3	.01	1	.21	9	.38
4	.09	4	.02	2	.21	28.0	.39
5	.08	5	.03	3	.22	1	.39
6	.08	6	.03	4	.22	2	.40
7	.07	7	.04	5	.23	3	.40
8	.07	8	.04	6	.23	4	.40
9	.06	9	.05	7	.24	5	.41
19.0	.05	21.0	.05	8	.24	6	.41
1	.05	1	.06	9	.24	7	.42
2	.04	2	.06	25.0	.25	8	.42
3	.03	3	.07	1	.26	9	.42
4	.03	4	.07	2	.26	29.0	.43
5	.02	5	.08	3	.27	1	.43
6	.02	6	.08	4	.27	2	.44
7	.01	7	.09	5	.27	3	.44
8	.01	8	.09	6	.28	4	.45
9	.00	9	.10	7	.28	5	.45
		22.0	.10	8	.29	6	.46
		1	.11	9	.29	7	.46
		2	.11	26.0	.30	8	.47
		3	.12	1	.30	9	.47
		4	.12	2	.30	30.0	.47
		5	.13	3	.31	1	.48
		6	.13	4	.31	2	.48
		7	.14	5	.32	3	.49
		8	.14	6	.32	4	.49
		9	.15	7	.33	5	.49
		23.0	.15	8	.33	6	.50
		1	.16	9	.33	7	.50
		2	.16	27.0	.34	8	.51
		3	.17	1	.35	9	.51
		4	.17	2	.35	31.0	.51
		5	.18	3	.35	1	.52
		6	.18	4	.36	2	.52
		7	.18	5	.36	3	.53

GIVEN PERCENTAGE	SUB- TRACT	GIVEN PERCENTAGE	SUB- TRACT	GIVEN PERCENTAGE	SUB- TRACT	GIVEN PERCENTAGE	SUB- TRACT
4	.53	6	.62	8	.71	9	.80
5	.54	7	.63	9	.72	38.0	.80
6	.54	8	.63	36.0	.72	1	.80
7	.54	9	.63	1	.73	2	.81
8	.55	34.0	.64	2	.73	3	.81
9	.55	1	.64	3	.73	4	.81
32.0	.56	2	.65	4	.74	5	.82
1	.56	3	.65	5	.74	6	.82
2	.57	4	.66	6	.75	7	.83
3	.57	5	.66	7	.75	8	.83
4	.57	6	.66	8	.75	9	.83
5	.58	7	.67	9	.76	39.0	.84
6	.58	8	.67	37.0	.76	1	.84
7	.58	9	.67	1	.77	2	.85
8	.59	35.0	.68	2	.77	3	.85
9	.59	1	.68	3	.77	4	.85
33.0	.60	2	.69	4	.78	5	.86
1	.60	3	.69	5	.78	6	.86
2	.61	4	.70	6	.78	7	.86
3	.61	5	.70	7	.79	8	.87
4	.61	6	.70	8	.79	9	.87
5	.62	7	.71				



## INDEX

- Algebra, tests in, 214-220.
- Ancient history scale, 225-226.
- Arithmetic, measurement of, 58; Courtis Tests in, series B, 59-67; nature of Courtis examples, 61; directions for giving Courtis Tests, 61-63; scoring Courtis test papers, 63-67; standard Courtis scores, 67; remedial instruction in, 71-74, 85, 88-90; reasoning tests, 75-78; Woody scales, 78-91; Boston research tests in fractions, 91-96; Cleveland survey tests, 96-101; Kansas diagnostic tests in, 101-105; the teacher's problem in, 105-106; the next step, 106; bibliography, 107-109.
- Average, 261.
- Ayres, Leonard P., work on spelling vocabulary, 6; spelling scale, 6-7, 9; writing scale, 28-35.
- Ballou, Frank W., referred to and quoted, 276.
- Bibliography, on spelling, 22; on handwriting, 56-57; on arithmetic, 107-109; on reading, 154-155; on English composition, 179-180; on drawing, 191; on history, 209-210; on geography, 210; on language, 210-211; on music, 211; on high school tests, 226-228; on measurement of general intelligence, 252-253; on statistical methods, 262; on teachers' use of standard tests, 277-278.
- Binet-Simon test, old form, 230; Stanford revision of, 250-252; comparison of results, 251; use of Binet-Simon test and group tests, 250.
- Blewett, Ben, referred to, 2, 4.
- Bobbitt, J. F., referred to, 1.
- Boston, use of Ayres spelling scale in, 10; spelling list, 16; research tests in fractions, 91-96; copying test, 179; tests in geography, 204-207.
- Breed and Frostic scale for measuring the general merit of English composition, 178.
- Brown's silent reading tests, 151; standard scores, 151.
- Buckingham, B. R., extension of Ayres spelling scale, 10; spelling scale, 11-13.
- Butte survey, referred to, 18, 48, 54.
- Childs, H. G., use of Thorndike drawing scale, 189.
- Cleveland survey, 8, 48, 50, 96-101, 269.
- Cody, Sherwin, commercial tests, 224.
- Commercial tests, 224.
- Composition, measurement of, 156-180; Nassau County scale, 157-161; scoring results, 161-162; using results, 162-166; Willing scale, 166-172; the teacher's own scale, 172; Gary scale, 172-178; Thorndike's extension of Hillegas scale, 178; Harvard-Newton scale, 178; Breed and Frostic scale, 178; Starch's punctuation scale, 179; Boston copying test, 179; bibliography, 179-180.
- Connersville course of study in elementary mathematics, referred to, 71.
- Correlation, 262.
- Counts, George S., referred to, 101.
- Courtis, S. A., referred to, 58; tests in arithmetic, 58-70; Silent Reading test No. 2, 126-133; aim, 126; description of, 127; giving test, 127; scoring results and computing scores, 127; class record sheet, 128; interpreting and using results, 130; standard scores, 130; remedial measures, 132.
- Deviation, 262.
- Diagnostic tests in mathematics, 221-222.
- Distribution, 260.
- Drawing, how now measured, 181; Thorndike scale in, 181-186; require-

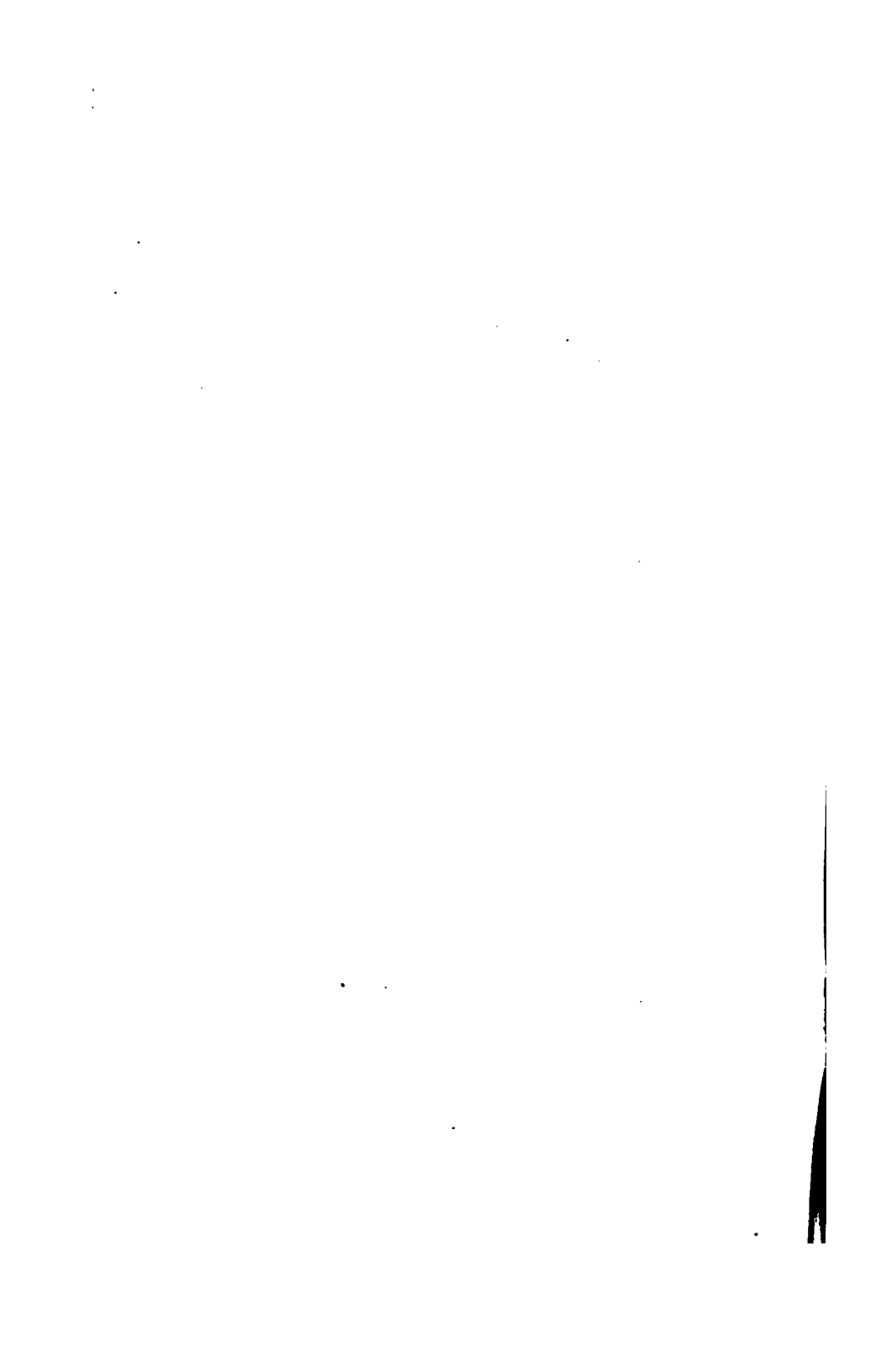
- ments of a scale, 183-184; how Thorndike scale was derived, 184-186; limitations of Thorndike scale, 186-188; grade standards, 188-190; using the scale, 190-191; bibliography, 191.
- Examinations, uniform, 267.
- Fordyce scale for measuring the achievement of reading, 149; standard scores in, 150.
- Freeman, Frank W., referred to, 1; standards in handwriting, 39, 40; analytical charts in handwriting, 43-44
- Geography, Starch test in, 202; Hahn-Lackey scale in, 202; Boston tests in, 204.
- Geometry, tests in, 220-221.
- Grading, present systems, 263-266; grading same paper, 266-267; using standard tests, 268-274.
- Grammar, tests in, 209.
- Grand Rapids survey, 97.
- Gray, C. Truman, score card for handwriting, 45.
- Gray, William S., oral reading test, 133-143.
- Haggerty, M. E., intelligence examinations: Delta 1 and Delta 2, 245-249.
- Haggerty and Noonan achievement examination in reading: Sigma 1, 153; standard scores, 153.
- Hahn-Lackey geography scale, 202.
- Handwriting, scale for, referred to, 1; measurement of, 23; first scale in, 24; Ayres scale, 24, 28-35; what to measure in, 24; giving the test in, 25-27; scoring for speed, 27; scoring for quality, 36; recording the scores, 37-38; standard scores, 38; standards for speed, 38-39; standards for quality, 40; social standard in, 41-42; remedial instruction, 42-47; Gray's score card, 45; proportion of children at standard quality, 48-50, 52-53, at standard speed, 49, 51; the Thorndike scale, 53-55; comparative scores, 55; Lister-Meyers scale, 56; bibliography, 56-57.
- Harvard-Newton scale for measuring English composition, 178.
- Henmon, V. A. C., Latin tests, 222-223.
- High School subjects, tests in, 212-228.
- Hiegar scale for measuring the quality of English composition, 156.
- History, measurement of, 192-202; Bell and McCullum test in, 193-195; Van Wagenen history scales, 196-201; diagnostic tests in, 201-202; bibliography, 209-210.
- Intelligence, general, measurement of, 229-234; differences among children, 233; Trabue language scales, 235-241; Otis group scale, 241-245; Haggerty's intelligence examinations, 245-249; Whipple group tests, 249; Binet-Simon test, Stanford revision, 250-252; bibliography, 252-253.
- Iowa, spelling scale, 11; elimination reports, 41.
- Iowa elimination reports, referred to, 41.
- Jones, W. Franklin, determines spelling vocabulary, 5; hundred spelling demons, 16-18.
- Kansas silent reading test, 149; standard scores, 149.
- Kelley, Truman L., history tests, 201.
- Language, Starch tests in, 207-208; Charters, test in, 208; Trabue, scales in, 235-241.
- Latin, tests in, 222-223.
- Median, 261.
- Middle fifty per cent, 261.
- Mode, 261.
- Montroe, Walter S., standardized silent reading test, 111-1; aim, 111; description of, 112; giving test, 114; interpreting and using results, 114; class record sheet, 115; standard scores, 116; scoring results, 114; remedial measures, 117; algebra tests, 215.
- Music, Seashore tests in, 209.
- Otis group intelligence scale, 241-245; aim, 242; description of, 242; giving test, 242; scoring results, 243; interpreting and using results, 244.
- Physics, tests in, 223.

- Quartile, 261.
- Reading, measurement of, 110-155; factors in, 111; Monroe's silent reading tests, 111-117; remedial measures, 117-118; Thorndike's scale for sentences and paragraphs, 119-126; Courtis silent reading test, 126-133; oral reading, 133; Gray's oral reading test, 133-143; Haggerty's visual vocabulary test, 143-148; Kansas silent reading test, 149; Fordyce scale, 149; Brown's silent reading tests, 151; Thorndike's scale for word knowledge, 152; achievement examination in, 153-154; bibliography, 154-155.
- Rice, J. M., referred to, 1, 20; spelling test, 13-16.
- Rogers, Anna L., diagnostic tests in mathematics, 221-222.
- Rugg and Clark, algebra test, 215-220.
- Sackett, L. W., ancient history scale, 225-226.
- Salt Lake City survey, 52, 53.
- Scales, stages of development, 3; uses of, 19-21; standard, 271-274.
- Seashore, C. E., music tests, 209.
- Silent reading test, 152.
- Smith, William Hawley, referred to, 20.
- Spelling, vocabulary for scale, 5-6; Ayres scale, 6-10; giving a test, 7; scoring the papers, 8; distributing scores, 9; Buckingham's extension of Ayres scale, 10; Iowa spelling scale, 11; Buckingham scale, 11-13; Starch test, 15-16; Boston minimum list, 16; Jones' demons, 16-18; pupil's list of misspelled words, 18; uses of spelling scale, 19-21; methods of teaching, 21; bibliography, 22.
- Springfield (Ill.) survey, 48.
- Starch, Daniel, spelling test, 15-16; standards in handwriting, 39, 40; punctuation scale, 179; geography test, 202; language tests, 207-208; physics test, 223; quoted, 266.
- Statistical methods, securing comparable results, 254-255; using a standardized test, 255-259; selecting the test, 256; scoring the papers, 257; tabulating results, 257; statistical terms, 259-262; standard tests, 268-274; applied to human product, 275-276.
- Stockard and Bell, geometry tests, 220-221.
- Stone, C. W., reasoning tests in arithmetic, 75-78.
- Surveys. See Butte, Cleveland, Grand Rapids, Salt Lake City, Springfield.
- Table of frequency, 260.
- Teachers' composition scale, 172-178; Gary composition scale as plan, 172; need of, 172; value to teachers, 178.
- Thorndike, E. L., referred to, 1, 24, 269; handwriting scale, 24, 53-55; reading scales, 119-126, 152; extension of Hillegas composition scale, 178; drawing, scale, 181-186.
- Trabue, M. R., Nassau County supplement to the Hillegas scale, 157-166; aim, 157; description of, 157; applying scale, 161; scoring results, 161; interpreting and using results, 162; class record sheet, 162.
- Visual vocabulary test, Part I, 143-148; aim, 143; description of, 143; giving test, 144; scoring results, 144; interpreting results, 145; using results, 146; corrective measures, 148.
- Visual vocabulary test, Part II, 151.
- Whipple group tests for grammar grades, 249.
- Willing scale for measuring written composition, 166-172; aim, 166; description and application of, 166; scoring results, 166; class record sheet, 167; interpreting and using results, 168; class scores, 168; scoring errors for general merit, 169.
- Withers, John W., referred to, 2.
- Woody, Clifford, arithmetic scales, 78-91.
- Zirbes, Laura, referred to, 270.











STANFORD UNIVERSITY LIBRARY  
LIBRARY, SCHOOL OF EDUCATION, STANFORD

To avoid fine, this book should be returned on  
or before the date last stamped below

MAY 15 1948

LIBRARY

371.27 .W748

C.1

How to measure,

Stanford University Libraries



3 6105 033 374 260

371.27

W748

cop. 1

276724

SCHOOL OF EDUCATION



